Molecular and ecological characterization of trypanosomatid parasites infecting honey bee

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

By

Qiushi Liu

Dec 2018

Table of contents

Abstract

The global decline of managed honey bee (*Apis mellifera*) colonies has stirred debate regarding the effects on worldwide crop production and ecosystem. Recent years, it has been shown *Lotmaria passim*, which is a newly classified honey bee trypanosomatid parasite, to be prevalent and associated with the weak colonies. Although there are many pathogens/parasites of honey bee, we still do not understand how *L. passim* changes the gene expression to adapt to the host environment and how the host responds to the *L. passim* infection by modifying the gene expression simultaneously. By using RNA-seq, we found that *L. passim* dynamically modifies the expression of mRNAs associated with respiratory chain and protein translation to adapt to the honey bee hindgut with anaerobic and poor nutritional conditions at the early stage and to become the dormant form at the late stage of infection. Meanwhile, some genes are continuously up- or down-regulated during the infection and they include GP63 as well as other proteins capable of modulating activities of host cell signalling pathways (up-regulated) and genes involved in detoxification of radical oxygen species. *L. passim* infection only slightly increases the honey bee mortality and does not affect the number of the gut microbiota; however, it induces the innate immune responses and the host appears to be under poor nutritional state by increase of mRNAs for *take-out* and *facilitated trehalose transporter* and decrease of *vitellogenin* mRNA. Meanwhile, pollen consumption can reduce the parasite load in the early stage of infection. At last, the CRISPR/Cas9 system was successfully applied to *L. passim* for the first time. Both exo- and endogenous genes were knocked out. These results help us to gain a deeper insight into the honey bee-trypanosomatid parasite interaction.

1. Chapter I

General introduction

1.1. Honey bee values and its gut trypanosomes

The honey bee (*Apis mellifera*) plays important roles in agricultural crop production and ecosystem conservation across the globe (Klein et al., 2007;Aizen et al., 2008;Potts et al., 2010). However, a decline in managed honey bee colonies has been observed in North America, Europe, and a part of Asia since 2006. The underlying reasons for the large-scale colony losses are complex but can be divided into several different categories: inadequate food supplies, anthropogenic chemicals, and exposure to various pathogens/parasites (Goulson et al., 2015). Honey bee-associated pathogens/parasites include fungi, bacteria, virus, mites and protozoans (Evans and Schwarz, 2011). Among the protozoans, there are two *Trypanosomatidae, Lotmaria passim* and *Crithidia mellificae*, which can infect honey bee. *C. mellificae* was first identified in Victoria and New South Wales, Australia in 1967. When this species was discovered, it was shown not to increase the mortality rate of honey bee (Langridge and McGhee, 1967). In the past 50 years, there has been almost no report on this species. In 2011, the prevalence of *C. mellificae* was found to peak in winter. Later, *C. mellificae* was reported to be a factor related to the colony losses when co-infected with microsporidia (*Nosema ceranae*) during the overwintering period in Belgium (Ravoet et al., 2013). In 2015, researchers compared the previous *C. mellificae* sample collected in 1967 to the two recent *C. mellificae* strains, San Francisco (SF) and Bee Research Lab (BRL). This comparison led to discover a novel honey bee infecting protozoan named *L. passim*. This is a new taxon which is distinct from *C. mellificae*. Specifically, as shown in Figure 1.1, it has a tear-drop shape with the elongated promastigote morphology that narrows posteriorly to a short caudate extension. *L. passim* promastigote has a broad, deep lateral groove. The parasites mainly colonize in the hindgut and become spheroids and adhere to the gut wall in a single layer. They were found especially in the rectal papillae as well as in the lower ileum (Schwarz et al., 2015). After the re-classification of this novel species, previous samples related to *C. mellificae* were re-analyzed and all of those samples were identified as *L. passim* (Ravoet et al., 2015;Cavigli et al., 2016). In fact, the earliest global finding of *L. passim* can be traced back to 2007 in Serbia (Stevanovic et al., 2016). After this re-classification, only samples from Belgium were confirmed to be *C. mellificae* at a low rate (Ravoet et al., 2015). Thus, whether *L. passim* or *C. mellificae* (together or separately) was related to the winter mortality should be reconsidered (Ravoet et al., 2013). Nowadays, *L. passim* appears to be the dominant trypanosomatid infecting *A. mellifera* all over the world in contrast to *C. mellificae* (SCHMID-HEMPEL and Tognazzo, 2010;Morimoto et al., 2013;Cepero et al., 2014;Huang et al., 2014;Ravoet et al., 2014;Cersini et al., 2015;Ravoet et al., 2015;Schwarz et al., 2015;Arismendi et al., 2016;Cavigli et al., 2016;Stevanovic et al., 2016;Vavilova et al., 2017;Castelli et al., 2018;Regan et al., 2018).

Figure 1.1 Scanning electron microscopic image of *L. passim*. It is a typical tear-drop shaped lanceolate promastigote cell with a deep, broad lateral groove. Scale bar: 4 µm (Schwarz et al., 2015).

1.2. Regulation of trypanosomatid gene expression, polycistronic transcription and trans-splicing

Next-generation sequencing (NGS) techniques make it possible for researchers to gain deeper insights into many organisms. Also, the decreasing cost enables NGS to be widely used for different experiments (Goodwin et al., 2016). Particularly, RNA-seq provides plenty of information regarding the levels of gene expression between different tissues, different conditions and different organisms at a high resolution. Generally, there are three aims for transcriptomics: 1) to identify all transcripts (including mRNAs, small RNAs and ncRNAs). 2) to reveal the transcript structures, such as the transcriptional start sites (TSSs), 5' and 3' ends and splicing patterns. 3) to quantify the expression levels of each transcript under different conditions. After obtaining the short reads, they are either aligned to a reference genome or assembled *de novo* without genomic information. As for the downstream analysis for RNA-seq data, differentially expressed genes (DEGs) analysis is the most common application of RNA-seq data. DGE is the genes differentially expressed across two or multiple conditions (Oshlack et al., 2010;Pertea et al., 2016). Due to the benefits that RNA-seq provided above, I used the top cited DGE analysis package, edgeR and the Trimmed Mean of M-values (TMM) normalization method embedded in this package to analyze the RNA-seq data for both *L. passim* and honey bee in order to understand the potential effects of *L. passim* infection on the honey bee.

Trypanosomatid parasites are unique organisms in many ways. They have special features on the genome organization and different mechanisms of gene expression and regulation which are distinct from other species. For example, trypanosomatid genome is organized into large polycistronic gene clusters (PGCs) in which many protein-coding genes are arranged

sequentially on the same strand of DNA. This was observed on chromosome 1 of *Leishmania major* (Myler et al., 1999). Later, with the help of NGS, similar genomic DNA organization was observed in *Trypanosoma brucei* (Berriman et al., 2005), *Trypanosoma cruzi* (El-Sayed et al., 2005) and the other chromosomes of *L. major* (Ivens et al., 2005). There are at least two PGCs on most of the chromosomes which can be transcribed either toward or away from telomeres (Martínez-Calvillo et al., 2010). However, genes within the same PGC do not have similar functions (Campbell et al., 2003). There are three different RNA polymerases (Pol): Pol I, II and III. Each Pol synthesizes specific RNA. For example, Pol I is responsible for transcribing ribosomal RNA including 18S, 5.8S and 28S rRNA. Pol II is involved in the production of messenger RNAs (mRNAs) as well as most small nuclear RNAs (snRNAs). Pol III generates essential RNA such as transfer RNAs (tRNAs) (Daniels et al., 2010;Martínez-Calvillo et al., 2010). For the processing of mRNAs in trypanosomatid, they undergo polycistronic transcription of protein-coding genes as well as *trans*-splicing (Johnson et al., 1987;Martínez-Calvillo et al., 2003). Mature mRNAs are produced by *trans*-splicing and polyadenylation from primary transcripts (Clayton, 2002). After the *trans*-splicing process, a Y-structure intermediate is generated instead of a lariat one which is different from *cis*-splicing (Michaeli, 2011). Genes within the same PGC have the same transcription level as a consequence of polycistronic transcription (Martínez-Calvillo et al., 2010). Thus, the gene expression level is determined by the post-transcriptional mechanisms in trypanosomatids (Teixeira, 1998;Clayton, 2002;Haile and Papadopoulou, 2007). However, detail of the post-transcriptional mechanisms to determine gene expression is not yet fully understood (Daniels et al., 2010). Previous research indicated the importance of 3' UTR of mRNA to regulate the gene expression in *Leishmania infantum* (Boucher et al., 2002) and *T. brucei* (Quijada et al., 2002). It has been recently shown that the codon choice plays the important role for regulation of mRNA levels in *T. brucei* (de Freitas Nascimento et al., 2018;Jeacock et al., 2018). Thus, gene expression in trypanosomatids is largely under the post-transcriptional control.

1.3. Introduction of CRISPR/Cas9 system

Recently, a new method for genome editing called clustered regularly interspaced short palindromic repeats (CRISPR) system has become widely used. CRISPR, together with the Cas9 protein, forms the CRISPR/Cas9 system. The CRISPR/Cas9 system has been applied for

genome manipulation in fruit fly, yeast, zebrafish, mouse and other species (DiCarlo et al., 2013;Gratz et al., 2013;Hwang et al., 2013;Wang et al., 2013). It has also been applied to edit the genomes of various trypanosomatid parasites: *T. cruzi* (Lander et al., 2015a;Peng et al., 2015;Lander et al., 2016b;Lander et al., 2017), *T. brucei* (Beneke et al., 2017;Rico et al., 2018), *L. major* (Sollelis et al., 2015b;Beneke et al., 2017), *Leishmania donovani* (Zhang and Matlashewski, 2015;Martel et al., 2017;Zhang et al., 2017) and *Leishmania mexicana* (Beneke et al., 2017). Although the nonhomologous end-joining (NHEJ) pathway appears to be absent in trypanosomatid parasites (Passos-Silva et al., 2010), the endogenous genes were successfully knocked out both by the microhomology-mediated end joining (MMEJ) and homology-directed repair (HDR) pathways, in order to repair Cas9-induced double-strand DNA breaks (DSBs). Generally, the CRISPR/Cas9 system can be divided into two parts, Cas9 protein and a single guide RNA (sgRNA). The CRISPR-associated protein, Cas9, is an endonuclease that uses a guide sequence to form base pairs with the target DNA sequence. This will enable Cas9 to introduce a site-specific double-strand break (DSB) in DNA. The sgRNA has two significant features. Firstly, a part of the sequence at the 5' end determines the target DNA site. Secondly, a duplex RNA structure at the 3' end binds to Cas9. This two-component system can target any given DNA sequences containing protospacer adjacent motif (PAM) by appropriately designing the sgRNA sequence, and thus it becomes a useful molecular tool in biology. A schematic depiction of the structure of CRISPR/Cas9 system is shown in Figure 1.2. The Cas9 enzyme (blue) generates breaks in double-stranded DNA using its catalytic centers (blades). These DNA cleavage sites are near PAM sequence (red). The 20-nucleotide sequence (orange) matches the target DNA sequence (gold). The green part binds and stabilizes the Cas9 protein. After the cleavage of DNA, the DSB can trigger repair enzymes to replace or disrupt DNA sequence at or near the cleavage site (Doudna and Charpentier, 2014). The establishment of CRISPR/Cas9 system in organisms needs both expression of sgRNA and Cas9. These two components can be expressed by different approaches as reviewed recently (Lander et al., 2016a).

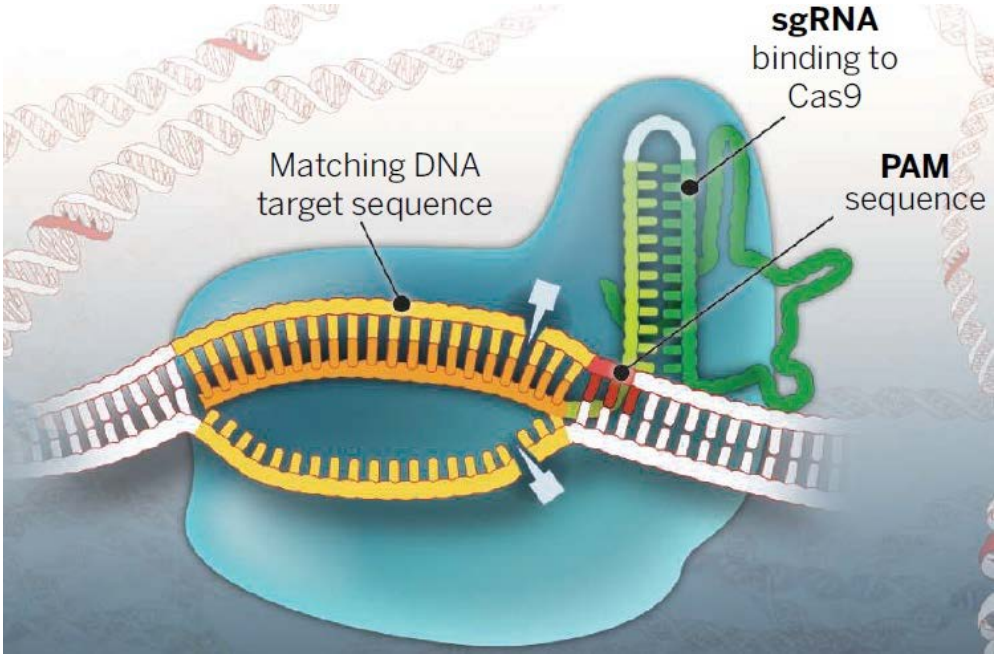


Figure 1.2 The structure of CRISPR/Cas9 system. The orange part represents 20 nucleotide of sgRNA which matches to DNA sequence of interest. The green part represents a loop which enables the sgRNA bind to Cas9 protein. Adapted and modified from (Doudna and Charpentier, 2014).

1.4. Aims of the project

All in all, the aims of my project are: 1) to understand the effect of *L. passim* infection on honey bee health under both natural and laboratory condition. 2) to identify the gene expression level differences associated with *L. passim* infection and analyze how infected honey bee response to infection by RNA-seq. 3) to establish the CRISPR/Cas9 genome engineering system in *L. passim*. The genes with dramatic changes of expression during the infection will be used as potential targets for CRISPR/Cas9 knock out.

2. Chapter II

Effects of *L. passim* infection

2.1. Introduction

2.1.1. Introduction of honey bee parasites and its global distribution

The global distribution of two honey bee trypanosomes (*C. mellificae* and *L. passim*) can be seen in Figure 2.1. By using species-specific PCR or real-time PCR, these two parasites can be detected according to their sizes without sequencing the amplicons (Szalanski et al., 2016;Xu et al., 2018). Interestingly, there are also several reports to indicate the potential interaction of *L. passim* with other pathogens. For example, a 9-year survey in Serbian honey bees indicates that *L. passim* tends to co-infected with *N. ceranae* (Stevanovic et al., 2016). The report from Chile also showed the same phenomenon as in Belgium that beekeepers suffered from winter losses and the reduction of colony activity with up to 90% of *L. passim* infection. However, only 18% of the hives analyzed were infected with *N. ceranae* (Arismendi et al., 2016). Another report using pooled honey bee samples indicates a positive relationship between *L. passim* and *N. ceranae* infection. Interestingly, the same report also demonstrated a negative correlation between the average monthly temperatures in the apiary and *L. passim* load, which may also imply that *L. passim* could contribute to the winter mortality (Vejnovic et al., 2018). However, another report showed there is no strong relationship between *L. passim* and *N. ceranae* infection (Tritschler et al., 2017). Last year, a study in India indicated *L. passim* could be co-infected with not only *N. ceranae*, but also *Nosema apis* for the first time (Vavilova et al., 2017). All of these results may indicate *L. passim* infection might be an important factor to honey bee loss since it may have interaction with other honey bee pathogens. Due to the new discovery and the high prevalence of *L. passim* infection, it is critical to understand the effects of *L. passim* infection on honey bee health.

The host-parasite interaction between bumble bee and its trypanosome, *Crithidia bombi* has been better understood compared with honey bee trypanosomes. In 1990s, researchers found *C. bombi* infection slows down the growth of bumble bee colonies in the early stage of colony establishment. Also, the size of ovary of the infected queen is smaller than that of the uninfected ones (Shykoff and Schmid-Hempel, 1991). Moreover, the infection of *C. bombi* modifies the bumble bees' behaviour and longevity under stress conditions (Brown et

al., 2003;Gegear et al., 2005). Furthermore, it has been shown that *C. bombi* infection on hibernated queen results in a loss of weight (Brown et al., 2003). In addition, *C. bombi* infection was also reported to impair the ability of bumble bee to utilize floral information (Gegear et al., 2006).

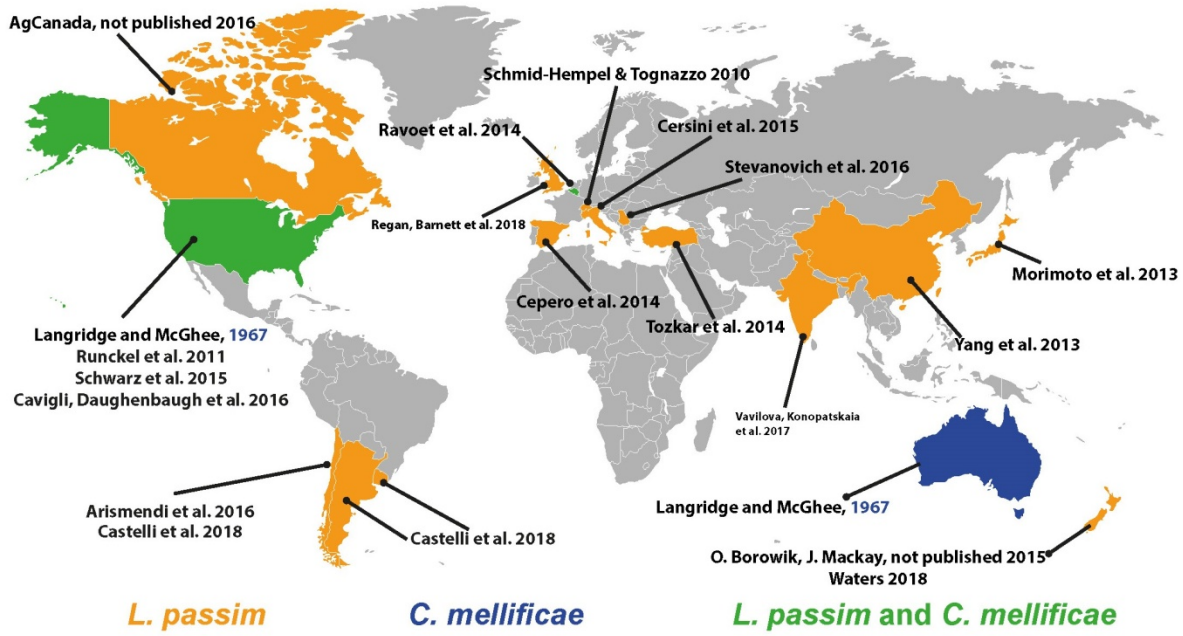


Figure 2.1 Global distribution of *L. passim* and *C. mellificae*. The orange parts indicate the existence of *L. passim* and the dark blue parts indicate the existence of *C. mellificae*. The green parts indicate both species were reported.

2.1.2. Bacteria species distribution in the honey bee gut

In newly-emerged honey bee workers, almost no bacterium can be found in the guts. However, the workers can obtain the core bacteria in 72 to 120 h after the emergence (Powell et al., 2014;Hroncova et al., 2015;Kwong and Moran, 2016). These core gut bacteria can be transmitted in three different ways: faecal-oral transmission, hive material transmission and oral trophallaxis (Kwong and Moran, 2016). Honey bee digestive tract can be divided into three major parts, which are foregut, midgut and hindgut. As shown in Figure 2.2, different bacterial species colonize in the different parts of the gut. Compared with foregut and midgut, honey bee hindgut has a higher number of bacteria with more diversity (Martinson et al., 2012). In the honey bee gut, the different parts have the specialized functions. For example, foregut functions for collecting nutrients from nectar. Although these nectars can be used as energy sources for bacteria, it is difficult for bacteria to colonize in the foregut. For the midgut, the primary function is for digestion. Several reports indicate

that honey bee midgut can absorb simple compounds from the pollen such as fructose, glucose and amino acids (Crailsheim, 1988;Crailsheim, 1990). The epithelial cells could secrete enzymes which can digest foods (Martinson et al., 2012). Meanwhile, the midgut epithelial cells also generate the peritrophic membrane (PM), which functions as a barrier for pathogens. The existence of PM prohibits the binding of bacteria (Tellam, 1996) and this may be the reason for relatively few bacteria present in the midgut. For the hindgut, it contains two parts, which are ileum and rectum. In ileum, several Gram-negative bacteria species colonize. *Snodgrassella alvi* forms a layer on the gut wall. On the top of this layer, *Gilliamella apicola* forms another layer (Martinson et al., 2012;Powell et al., 2014). Rectum, which is the last region of the gut, may function for re-absorption of salts and water before defecation. In the rectum, the majority of bacteria are related to fermentative Gram-positive species such as *Lactobacillus* Firm-5, Firm-4 and *Bifidobacterium asteroides*. The rectum contains empty pollens which cannot be used by honey bee (Roulston and Cane, 2000). However, these exines can be utilized by bacteria (Warnecke et al., 2007) and the rectum harbours the majority of bacteria in the honey bee gut (Martinson et al., 2012).

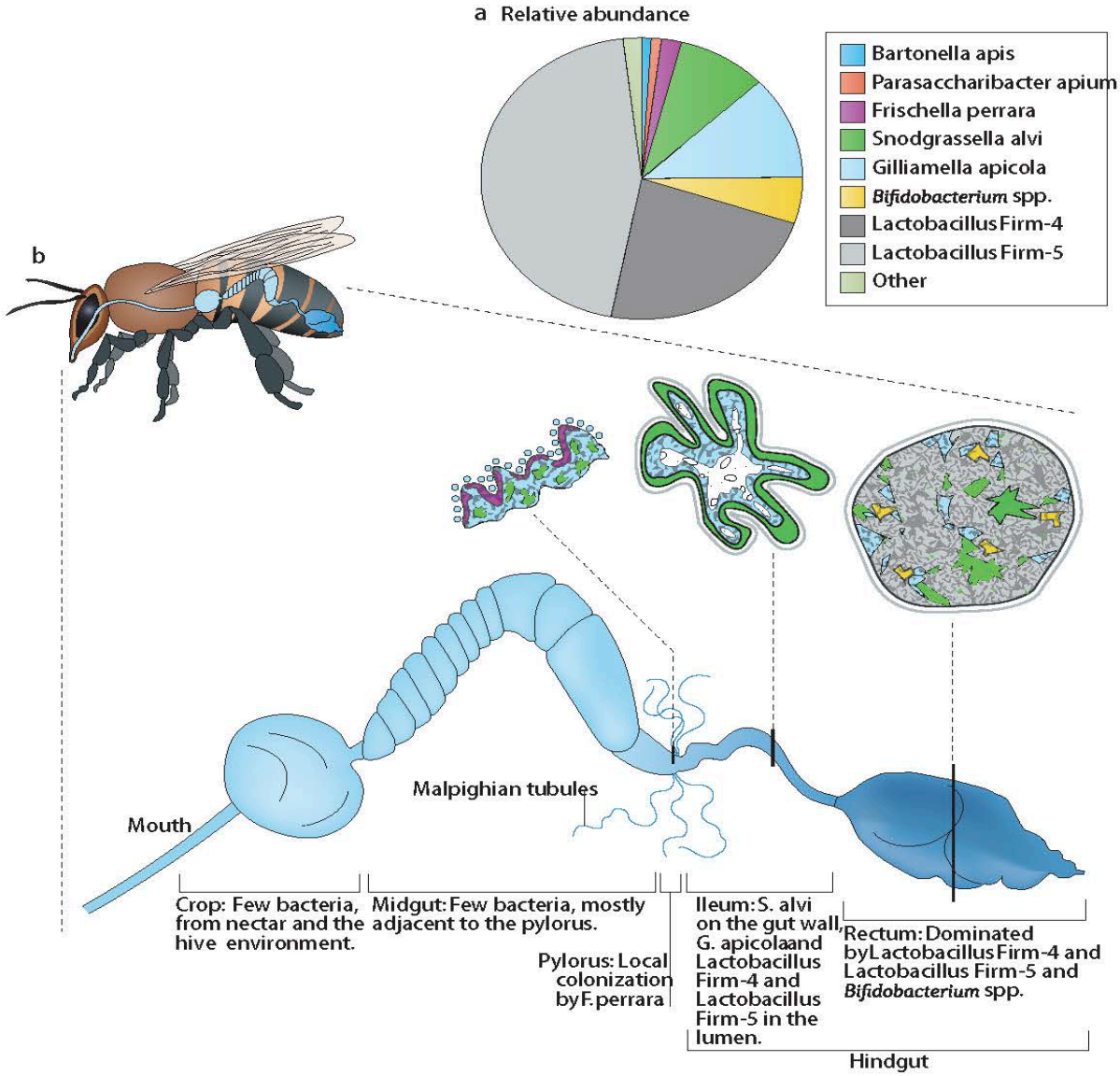


Figure 2.2 Relative abundance of bacteria in the different parts of honey bee gut. a) Relative abundance of bacteria in honey bee worker gut based on 16S rRNA sequencing (Moran et al., 2012). b) Different bacteria in the different locations of the gut (Martinson et al., 2012;Powell et al., 2014). Adapted from (Kwong and Moran, 2016).

2.1.3. Functions of core bacterial communities in the honey bee gut

From the previous studies, bacteria present in honey bee gut can be divided into nine distinct phylogenetic lineages (Martinson et al., 2011). The presence of different bacterial communities in the honey bee gut has been proposed as a consequence of long-term co-evolution process. However, it should be pointed out that these gut bacteria are not always beneficial to their host. For example, *Frischella perrara*, which is a core bacterium near the honey bee pylorus can stimulate the immune response such as melanization, which is always

associated with the invasion of pathogens (Engel et al., 2015). Another critical role of gut bacterial communities is for neutralizing toxic substances from food. For example, proteins from specific *Gilliamella apicola* stains can digest the cell wall of pollen. This step is essential for honey bee since pollen is a critical part of their daily food consumption. Honey bee may have to utilize bacteria to digest pollen in order to meet the food requirement. More recently, *in vivo* experiments show that honey bee gut bacteria help young adult workers to gain weight, increase the expression of *Vitellogenin* (*Vg*) and increase the sensitivity of sucrose. Moreover, these gut bacteria also shape the gut microenvironment not only lowering the oxygen but also the pH level. Specifically, the bacteria near the gut peripheral areas consumed the oxygen so that the inner part of the gut maintained anoxia which is required for bacterial activities. Furthermore, these bacteria also produce some short-chain fatty acid which results in the generation of metabolites such as acetate (Zheng et al., 2017). *Lactobacillus*, *Acetobacteraceae* and *Bifidobacterium* can also produce short-chain fatty acids during carbohydrate metabolism. Absorption of these fatty acids can supply energy for honey bee (Anderson et al., 2011;Martinson et al., 2011). Genomic study also indicates these bacteria communities also involved in the pathways of breakdown carbohydrates (Engel et al., 2012;Lee et al., 2015). These bacteria include the genera *Lactobacillus* and *Bifidobacterium*. Specifically, *B. asteroides* species clade was shown to express specific genes which participate in carbohydrate utilization (Bottacini et al., 2012). Another study indicates the significant variation of carbohydrate breakdown genes in *Bifidobacterium* spp. (Ellegaard et al., 2015). As for *Lactobacillus* strains, they were clustered into two different phylogenetic clades in the honey bee gut, which are Firm-4 and Firm-5. These two bacterial clusters were shown to have a large number of genes related to the uptake of sugar. *B. asteroids* and Firm-5 were also shown to have potential roles in the synthesis and degradation of trehalose, which is used as energy storage in insects (Ellegaard et al., 2015). These core bacterial communities also play critical roles in carbohydrate and fatty acid metabolism. More recently, by using metabolomics, it has been proposed that the honey bee gut bacterial communities can utilize the indigestible compounds which from the daily diet (Kešnerová et al., 2017). These include the components of outer pollen wall and flavonoids. Ultimately, these components were fermented into organic acids by gut bacteria which may influence the gut microenvironment (Kešnerová et al., 2017). Interestingly, the cross-feeding phenomenon was also shown within different gut bacterial communities, i.e. one species

can provide metabolites which can be utilized by other species. Specifically, different metabolites can accumulate during the growth of *G. apicola*. Also, different metabolites were used during *S. alvi* growth which indicates a cross-feeding relationship within gut bacterial communities (Kešnerová et al., 2017). All these results indicate the critical function of honey bee gut bacteria.

2.1.4. Interactions between bacterial communities, parasites and the host

These bacterial communities are supposed to have different functions for the honey bee host. For example, using antibiotics to eliminate core bacteria in the bumble bee gut results in an increased susceptibility of *Crithidia* infection compared with untreated bees (Koch and Schmid-Hempel, 2011). Another indirect effect was reported with wild bumble bee in which *Crithidia* infection level was negatively associated with the relative abundance of *Gilliamella*, a core bacterium in the gut (Cariveau et al., 2014). Moreover, the composition and diversity of bacteria can also influence the bumble bee health (Koch and Schmid-Hempel, 2012). *In vitro* experiments also indicate that *Lactobacillus apis*, isolated from honey bee gut can inhibit the growth of other honey bee pathogens (Killer et al., 2014). All of these reports indicate the potential defensive role of core gut bacterial communities against other pathogens/parasites for the host. There are several reports directly or indirectly demonstrating the potential roles of bacterial communities for parasites. More recently, *Apibacter*, *Lactobacillus* Firm-5 and *Gilliamella* bacteria have been shown to correlate with lower *C. bombi* infection in the bumble bee gut (Mockler et al., 2018). *S. alvi* is one of the core residents in the honey bee gut (Jeyaprakash et al., 2003;Kwong and Moran, 2013;2015;Kwong and Moran, 2016). Recent study on the interaction between *L. passim* and *S. alvi* indicates that pretreatment with *S. alvi* makes normal bees more susceptible to *L. passim* infection under natural condition. Under stressed condition (no protein provided in a cage), the bees were more susceptible to *L. passim* infection and pretreatment of *S. alvi* did not change the susceptibility to *L. passim* infection. This result demonstrates the pre-inoculation of honey bee core bacteria unexpectedly has negative effects which lead to the gut dysbiosis (Schmidt and Engel, 2016). When the gut core bacterial communities were eliminated by antibiotics, honey bee was shown to increase susceptibility to *N. ceranae*

infection (Li et al., 2017). Similarly, tetracycline treated honey bees were shown to increase susceptibility to *Serritia* infection which also leads to increased mortality (Raymann et al., 2017). Interestingly, a study on the relationship between *L. mexicana* and *Lutzomyia longipalpis* illustrated the *Lu. longipalpis* can survive longer when *L. mexicana* exists with the pathogenic bacteria *Serratia marcescens* (Sant'Anna et al., 2014). This indicates *L. mexicana* can protect the sand fly host from another bacterial pathogen. The bacteria composition is also variable between the different castes and sex. Honey bee queen harbors different bacteria composition compared to workers and several particular species are missing (Tarpy et al., 2015). As for drone, although the majority of gut bacteria are similar to the worker, they were reported to have more *Lactobacillus* spp. and more variable species (Kapheim et al., 2015). Moreover, the previous report also suggested that the bacterial composition shifts toward to Gamma-1 and Gamma-2 proportions in the elder honey bee workers (Martinson et al., 2012). Another report in 2017 showed that pretreatment of the frames with *Lactobacilli* and *Bifidobacteria* can dramatically benefit the hive. Specifically, the bacterial treatment can increase the production of honey and pollen. The brood population was also high compared to the control bee hive (Alberoni et al., 2018). At last, gut symbiont, *B. asteroides*, can promote the production of honey bee hormones which is known to affect host development (Kešnerová et al., 2017).

L. passim is confined to the hindgut and stay together with the gut microbiota of *A. mellifera* worker (Martinson et al., 2011;Schwarz et al., 2015). In order to investigate the possible relationship between *L. passim* and different bacteria species inside the hindgut, I intend to divide honey bees into three groups and feed with different food combinations. Then, the relative abundance of universal bacteria, Firmicutes bacteria as well as *L. passim*, will be quantified by qPCR. The bacteria changes by *L. passim* infection may alter the host weight, metabolism pathways and other host hormones which may affect the bee fitness. Honey bee Vg is a female-specific protein synthesized in the fat body of the abdomen with the molecular weight of 180 kDa (Wheeler and Kawooya, 1990). Honey bee Vg is released into the haemolymph and it can be transported to other tissues (Amdam et al., 2003). Vg functions in different ways for helping to establish the honey bee social organization on the division of labour and foraging specialization (Amdam and Omholt, 2003;Nelson et al., 2007). Moreover, Vg expression is also related to the resistance to oxidative stress. Specifically, the previous study proposed the expression of Vg can extend the lifespan of workers and queen

(Seehuus et al., 2006;Corona et al., 2007;Nelson et al., 2007). Vg is also a nutrient-sensitive protein involved in feeding behaviours. The levels of *Vg* mRNA are correlated with the state of adult nutrition (Di Pasquale et al., 2013). Furthermore, Vg is also an effective regulator of the immune response of honey bees (Buttstedt et al., 2014). There are several reports to indicate how honey bee pathogens can affect the expression of Vg. For example, previous report elucidated that the Vg expression was slightly suppressed compared with control bees when inoculated with the core gut bacterium *S. alvi* and *L. passim* in the hive (Schwarz et al., 2016). Similarly, not only with endoparasites, honey bee infected by ectoparasite such as *Varroa destructor* was also reported to suppress the expression of Vg (Amdam et al., 2004). Thus, Vg is a critical indicator of honey bee fitness. Thus the relative abundance of honey bee *Vg* in the fat body was also tested in the *L. passim* infected bees. The Vg expression in the fat body altered by *L. passim* infection may alter the nutritional states such as protein and fatty acid metabolism. Moreover, the altered Vg expression may also affect the behavior and accelerate senescence. Not only on individual level, *L. passim* might be transmitted via faeces from infected bees to other nest mates by faecal-oral route or by visiting the same flower. Furthermore, the weight of *L. passim* infected bees and sham treated bees were measured in order to understand the sub-lethal effect of *L. passim* infection.

2.2. Materials and Methods

2.2.1. Honey bees and *L. passim*.

A. mellifera hives were bought from a honey bee keeper in Jiangsu Province, China. The honey bee colonies are maintained at XJTLU campus and can feed *ab libitum* from April to November. From December to March, pollen mixed with 50% (v/v) sucrose is added to the honey bee hives in order to provide sufficient food. *L. passim* strain PRA-403 isolated from *A. mellifera* gut in San Francisco (strain SF), the US in 2010 was bought from the American Type Culture Collection (ATCC).

2.2.2. Preparation for culture medium of *L. passim*

The culture medium of *L. passim* was prepared by modifying the culture medium for *C. bombi* (Salathe et al., 2012) and is shown below.

Components of culture medium for *L. passim*.

FP stock solution[1]

NaCl	2.8 g
KCl	0.4 g
Na_2HPO_4	8.875 g
Tryptose Soya Broth (BIODEE)	10 g
Brain Heart Infusion (HANGWEI)	2 g

Ad ddH_2O to the total volume of 1000 mL; use HCl set to pH of 5.8

10×FB stock solution[2]

D(-)Fructose (BIODEE)	1.8 g
L-Proline (BIODEE)	0.289 g
Vitamine B1 hydrochloride (BIODEE)	0.0067 g
Folic acid (Biosharp)	0.0044 g

Ad FP stock solution to fill to 100 mL

Haemin stock (2 mg/ml)[3]

Haemin-chloride (BIODEE)	100 mg
NaOH (1 M)	1.25 mL
ddH_2O	23.75 mL

Dissolve by heating to boiling point, cool down and add H_2O to a final volume of 50 mL

Heat inactivated Fetal Bovine Serum (hiFBS) (BOVOGEN)

Full Medium[4]

FP stock solution	36 mL
10×FB stock solution	4 mL
Haemin (BIODEE)	44.5 µL
hiFBS	4.5 mL
100X Penicillin-Streptomycin (Beyotime)	450 µL
Gentamycin (NobleRyder)	22.5 µL

[1]Autodave and store at 4 °C.

[(2)]Filter through 0.22 µm filter and store at -20 °C.

[(3)]Filter through 0.22 µm filter and store at -20 °C.

[(4)]Store at -20 °C.

2.2.3. Inoculation of honey bees with *L. passim* and dissection of the guts

To prepare *L. passim* cells for inoculations, promastigotes of PRA-403 were grown in modified FPFB medium supplemented with 100X penicillin-streptomycin, and 50 mg/mL gentamycin (NobleRyder) at 25 °C. After 48-72 h, cultures were pelleted at 2,000 x g for 5 min at room temperature. The medium was discarded and the pellets were washed once with PBS followed by resuspension in 10% sterile sugar water (1:1 v/v) and 90% PBS at 20, 000 viable cells/µL. Broods with capped cells were collected and incubated at 33 °C. A basin was put in the incubator to maintain the relative humidity in the incubator as described (Williams et al., 2013). New emerged adult honey bee workers were collected to a plastic box and fasted for 2-3 h. A pre-inoculation experiment was conducted before inoculation to guarantee the majority of the newly emerged workers are willing to intake the inoculums. The new emerged honey bee workers were divided into two treatment groups: One group was hand fed with 5 µL of 50% (v/v) sucrose solution. The other group was fed with 5 µL of sterilized sucrose solution containing *L. passim* promastigotes (100,000 cells in total). The honey bees were then put back to the hive and fed *ad libitum*. The honey bees were sacrificed and dissected at day 3, 11, 19, 25, 32 and 37 post infection (PI) under a dissection microscope. Gut tissues were collected from the honey bees and immersed in 30 µL PBS on a glass slide and examined with a DIC microscope at 200×, 400×, 600× and 1,000x magnification.

2.2.4. Quantitative PCR (qPCR) to analyze the abundance of *L. passim* in the infected honey bee guts

New emerged honey bee workers were maintained and obtained as described above. Similarly, *L. passim* promastigotes and infection procedures were followed as described above. All the bees were kept in metal cages at 33 °C under the laboratory condition. 10 live bees from five different time points (day 1, 3, 8, 15 and 22 PI) were sacrificed and the gDNA

from the whole abdomen was isolated individually. 1 µL of gDNA was used as a template for qPCR reactions. SYBR Green I (Takara) together with ROX reference dye II was used in the qPCR reaction. The qPCR primer sets to detect both *L. passim* and honey bee were designed using Primer Prime 5 software and NCBI online primer design software. For *L. passim*, *internal transcript spacer 2 (ITS2)* ribosomal RNA sequence was used for the primer design. The *AmHsTRPA* gene of honey bee was used as an internal control gene. Specific primers for qPCR can be viewed in Appendix 1. The cycling conditions are as follows: for holding stage, 95°C for 30 sec; cycling stage (40 cycles), 95°C for 3 sec, 60°C for 30 sec; melting curve: 95°C for 15 sec, 60°C for 60 sec and 90°C for 30 sec. The qPCR results were analyzed by Steel-Dwass (Dwass, 1960;Steel and Torrie, 1960) method to illustrate the differences between groups.

2.2.5. Mortality test for *L. passim* infected honey bee workers

In total, 300 infected bees and 300 sham treated honey bees workers were incubated in metal cages at 33 °C. Similarly, 240 infected bees and 240 sham treated honey bee workers were incubated in metal cages at 25 °C. The metal cages were put separately to prevent the potential trophallaxis between groups. The infection and sham-treated protocols were used as described in section 2.2.3. Cotton containing 50% (v/v) sucrose were put at the bottom of the metal cages to provide sufficient foods for the bees to prevent death caused by starvation. Cotton was replaced every 7 days. All the mortality tests from different temperatures were from 3 individual experiments. In the first experiment at 33 °C, 3 dead bees from infected group (15 days post-infection), 3 from 16 days post infection together with 1 dead bee from the control group were homogenised using DNAzol® reagent (Thermo Fisher) to extract the genomic DNA. *L. passim* ITS1 region was used to detect the existence of parasite and the honey bee *AmHsTRPA* was used as control (Da Silva et al., 2004). The primer sets were used in the Appendix 1. The graphs were performed by GraphPad Prism 6.05 software. The data were analyzed by Log-rank (Mantel-Cox) test to determine the differences between different survival curves.

2.2.6. qRT-PCR analysis for honey bee fat bodies *Vg* mRNA in the *L. passim* infected honey bee and the control bees

Similarly, as in the infection experiment, honey bees were divided into two groups: *L. passim*

infected and sham treated. The amount of *L. passim* promastigotes was fed as in the infection experiment as well as the sham treated group. All the bees were marked on the thorax and released back to the original hives. 20 days PI, bees from both groups were collected back to the lab for dissection. Fat body tissues were collected from both groups followed by RNA isolation using TRIzol® (Thermo Fisher). For reverse transcription, 1 µL total RNA was reverse transcribed using random hexamer primers with RTace reverse transcription enzyme from Takara. The complementary DNA (cDNA) library was used as templates for qPCR. The primers for honey bee *Vg* was used from the previous paper (Ihle et al., 2015). Honey bee 18s ribosomal RNA primers were used for endogenous control which is shown in the Appendix 1. The primer efficiency was calculated first from the serial dilution of the cDNA libraries.

2.2.7. qPCR analysis for relative abundance for universal bacteria, *Firmicutes* and *L. passim* under different conditions

In order to understand the interaction between *L. passim* and the bacteria, different infection experiments were performed as shown in Figure 2.3. Before treatment, newly emerged honey bee workers were exposed on the frames for 48 h in order to receive the core bacteria communities in the guts. Honey bees were divided into three groups for different treatment and food provided. Specifically, honey bees from group A were initially fed with sucrose only. Group B and C were fed with *L. passim*. Later on, bees were put back into cages according to the groups and fed with different food. In group A and B, bees were fed with sucrose and pollen. In group C, bees were fed with only sucrose. All the bee cages were maintained under the laboratory conditions as shown above. There are three questions addressed according to the experimental design. The comparison between group A and group B could answer the question: does *L. passim* infection alter *Firmicutes* and universal bacteria communities? The comparison between group B and C could answer the question: does pollen alter the *L. passim* load in the infected honey bee guts? *L. passim* ITS2 ribosomal RNA primer was used to detect the relative abundance of *L. passim* which can be seen in the Appendix 1. Primers for universal bacteria and phylum-specific *Firmicutes* were adapted from previous papers (De Gregoris et al., 2011;Powell et al., 2014). The honey bee *beta-actin* gene was used as a reference gene (Li et al., 2017). The efficiency of primers was

determined and the relative expressions of each target were calculated by using ΔC_t. As for analysis, normalized ΔC_t was used for the unpaired t-test. Two-tailed calculation together with the statistical significance ($p < 0.05$) was used for the statistical analysis at each time points. The morphology of the mixed pollen can be seen under the light microscope in Figure 2.4.

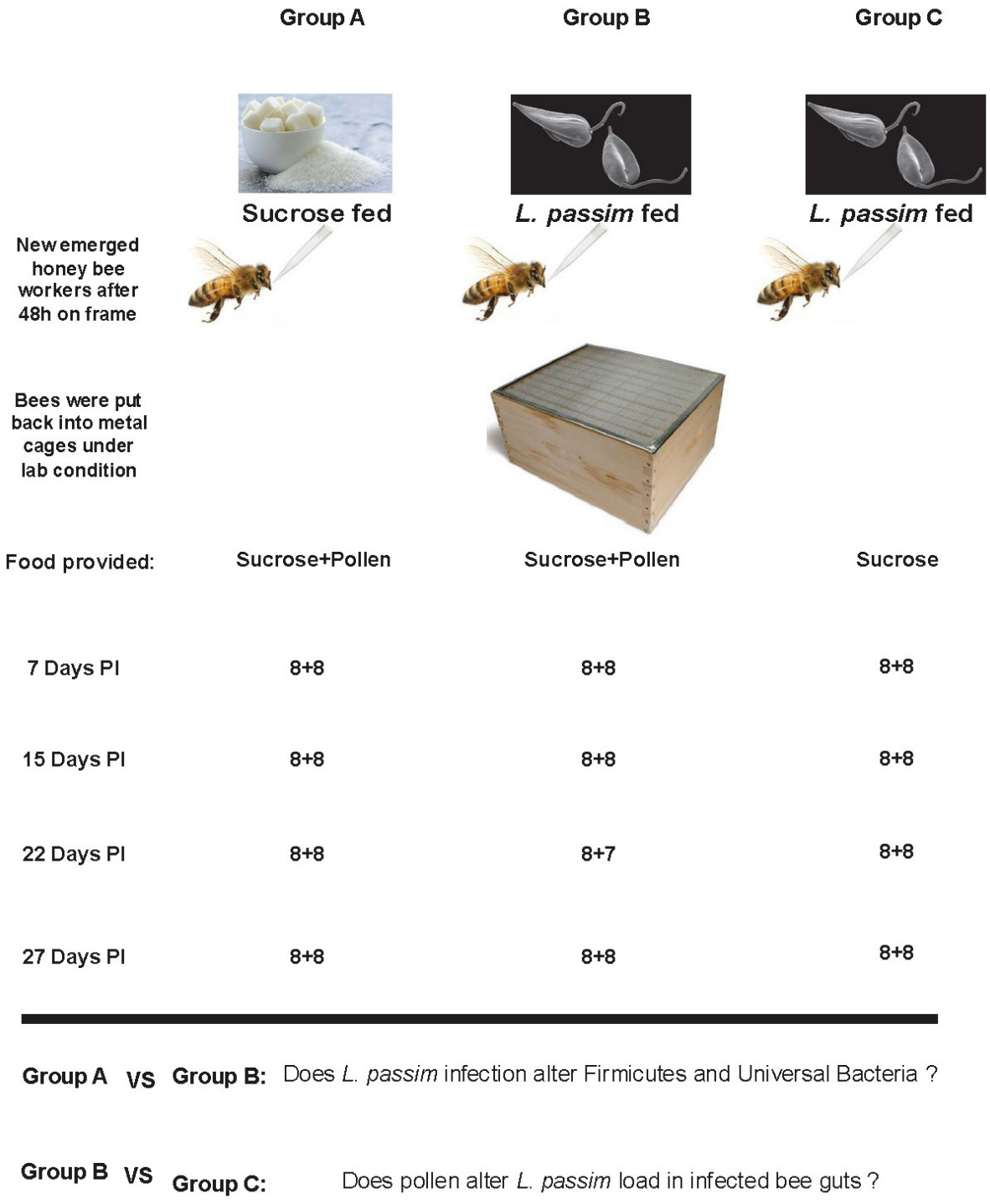


Figure 2.3 Schematic representation to measure the relative abundance of *Firmicutes*, universal bacteria and *L. passim* load under different conditions. Group A was fed with sucrose. Group B and C were fed with *L. passim*. Subsequently, Group A and B were fed with sucrose and pollen. Group C fed with sucrose only. The number of bees from each group

with different treatments was shown in the figure. Two questions were addressed at the bottom by comparing the different groups.

Figure 2.4 Morphology of the mixed pollen can be seen under light microscope. There are at least two distinct pollen species present in the pollen mixture. Magnification: 400X; Bar=50 µm.

2.2.8. Honey bee weight measurement in *L. passim* infected group and control group

L. passim infected and sham treated bees were performed as described in 2.2.3. Bees were marked on the thorax with different colours and put back to the hive. 20 days PI, bees from different groups were collected back to the lab and immobilized on ice. Then, bees were measured on the electronic balance individually. In total, two individual experiments were performed in duplicate and the statistical analysis was performed individually. Different hives were used in the individual experiments. In the first experiment, 27 bees from each group were collected for the weight measurement. In the second experiment, 16 bees from infected group and 15 bees from sham treated group were used for weight measurement. Similarly, unpaired t-test with two-tailed calculation was performed. The statistical significance was set to 0.05. All statistical analyses were performed by GraphPad Prism 6.05.

2.3. Results

2.3.1. Infection of *L. passim* to honey bee under natural and laboratory conditions

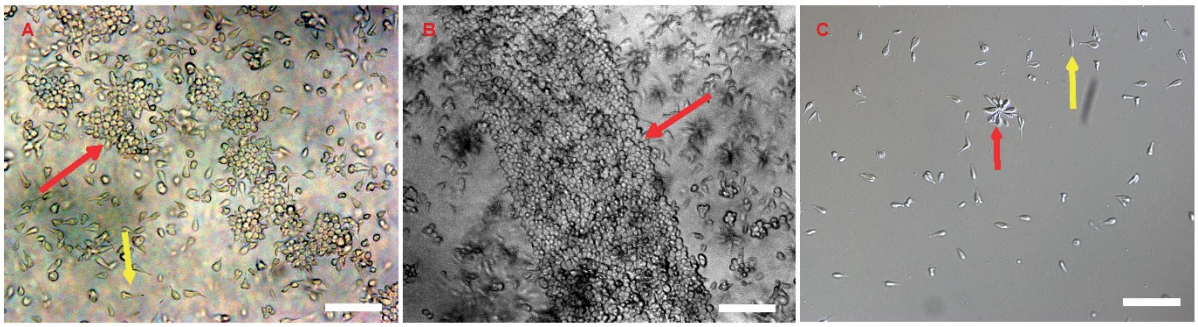


Figure 2.5 *L. passim* with different morphology can be observed during culture in the flask. A) *L. passim* in the modified FPFB medium. The red arrow illustrates the parasites can accumulate into a small cluster and attach to the flask. The yellow arrow illustrates the free-swimming form of the parasite. 400X. Bar=20 µm. B) *L. passim* in the modified FPFB medium can also accumulate as a large cluster with the linear boundary (red arrow) at the bottom of flask if they are cultured for long period (e.g. more than 15 days). 400X. Bar=20 µm. C) The red arrow demonstrates that *L. passim* can form a rosette in the flask which may indicate the rapid growth in the medium. The yellow arrow demonstrates the free-swimming ones in the flask. Differential interference contrast (DIC) microscopy. 400X. Bar= 40µm.

In Figure 2.5, two forms of *L. passim* with different morphologies can be found during culture with the liquid medium. The attached form of *L. passim* appears to be similar to that in the infected honey bee gut. As a member of *Leishmaniinae* (Schwarz et al., 2015), *L. passim* can also form a rosette in the medium which may indicate the rapid cell division in the medium or even a specific stage during their life cycle (Iovannisci et al., 2010).

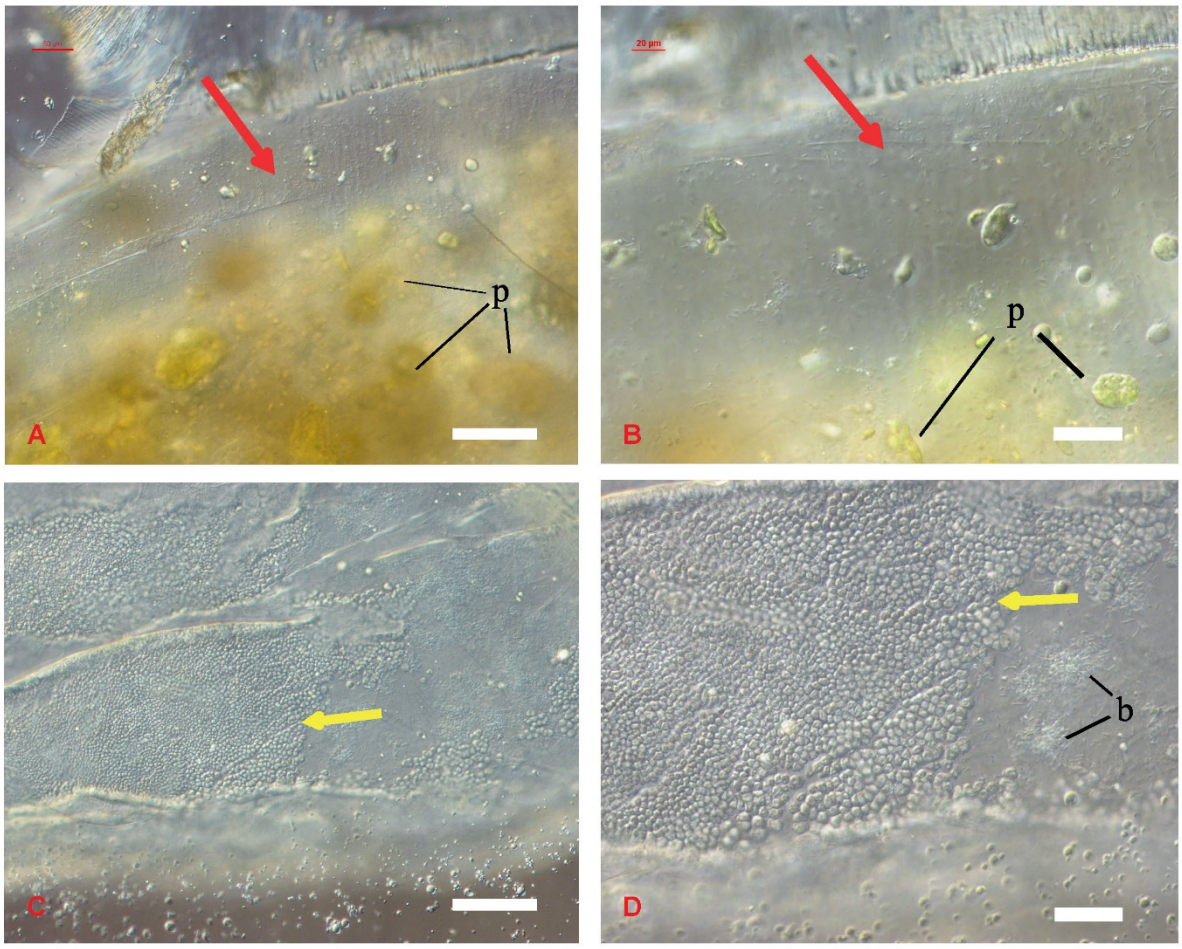


Figure 2.6 Honey bee hindgut at 19 days after the infection of *L. passim* under natural condition. (A and B) The hindgut of sham treated bee (uninfected control). The red arrows show the smooth gut epithelium with no parasites attached to it. (C and D) The hindgut of *L. passim* infected bee. The yellow arrows show that *L. passim* forms a single dense layer of spheroids along the luminal surface. p: pollen grains; b: unidentified bacteria. Magnification: A and C, 200x, bar=200 µm; B and D, 400×, bar=40 µm.

Figure 2.6 illustrates a significant difference between the guts of *L. passim* infected and the sham treated bees at 19 days after the infection. Specifically, in the sham treated group, the luminal surface is smooth with no parasites attached to it. However, for the infected group, *L. passim* forms a single dense layer along the luminal surface.

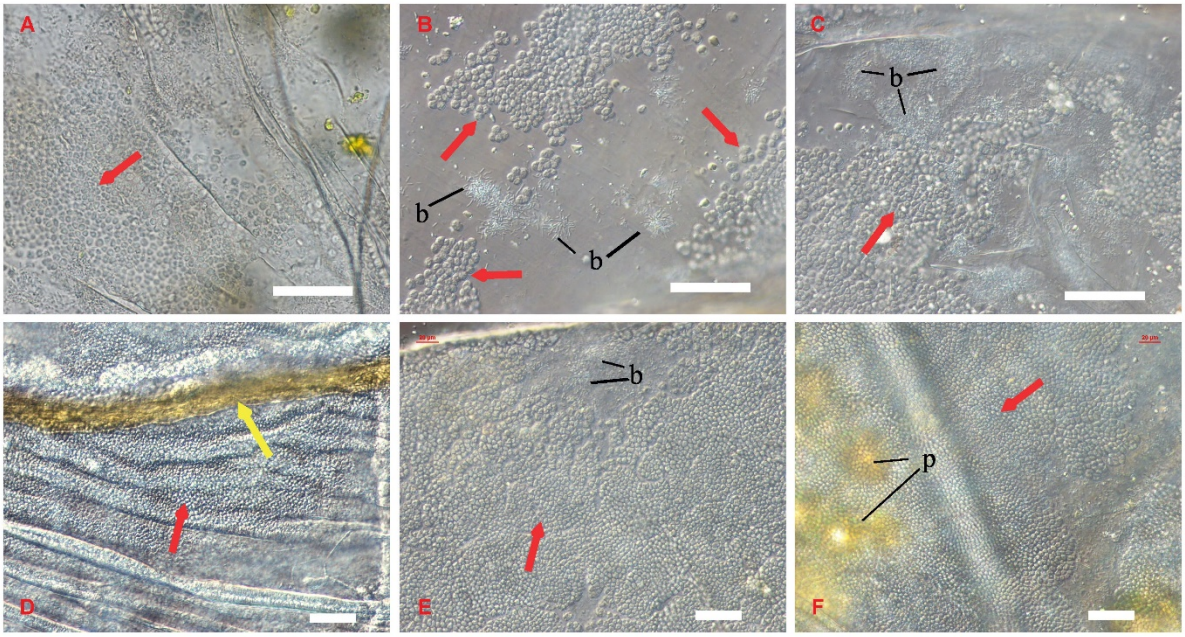


Figure 2.7 *L. passim* infected honey bee hindgut at day 3 (A), 11 (B), 19 (C), 25 (D), 32 (E), and 37 (F) under natural condition. Yellow arrows show the rectal pads of *A. mellifera*. Red arrows demonstrate that *L. passim* forms the single dense layer of spheroids along the luminal surface. p: pollen grains; b: unidentified bacteria. Magnification: A, B and C: 600x; D, E and F: 400×, bar=40 µm.

Figure 2.7 shows the *L. passim*-infected honey bee guts at 3, 11, 19, 25, 32, and 37 days after the infection. It takes less than 3 days for *L. passim* to migrate to the hindgut since the accumulated form of parasites can be seen in the hindgut at 3 days after the oral infection. Afterwards, *L. passim* attaches to the surface of hindgut and forms a single layer of spheroids. Also, the number of parasites seems to increase by time after the infection. Later, the hindgut surface was completely occupied by the parasites which remain the same afterwards (consistent with the qPCR results).

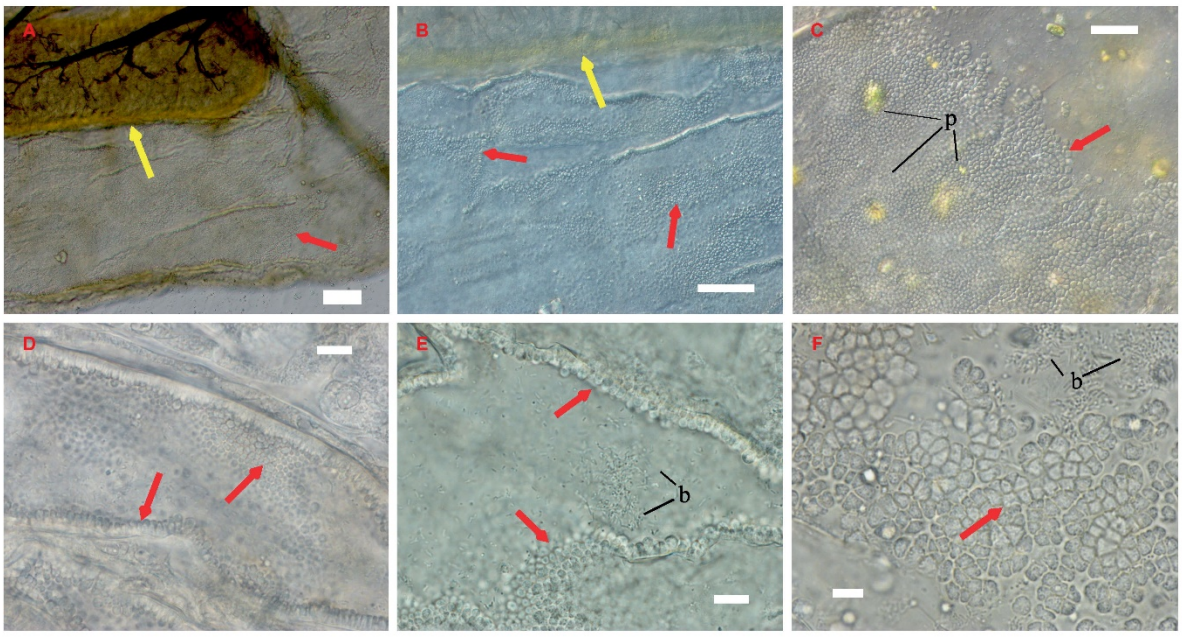


Figure 2.8 *L. passim* infected honey bee hindgut at 19 days after the infection observed at different magnifications. (A, B and C) The red arrows show the positions of *L. passim*. The yellow arrows show the rectal pads of the honey bee. Magnification: 100X, 200X and 400X. Bar=200 µm, 100 µm and 40 µm. (D and E) Red arrows illustrate that *L. passim* attaches along the luminal surfaces with a high density. Unidentified bacteria can also be found in the hindgut lumen (b).Magnification: 600X oil lens. Bar=20 µm. F) *L. passim* attaches to the hindgut epithelium along with unidentified bacteria (b). Magnification: 1,000X oil lens. Bar=10 µm. p: pollen gains.

Figure 2.8 demonstrates the overall location of the parasites in the honey bee hindgut. Specifically, the parasites accumulated near the rectal pad as described previously (Schwarz et al., 2015). Unidentified bacteria can also be found in the lumen of hindgut together with parasites. Occasionally, several drops of faeces were found from the infected honey bees (25 days PI) when they were put in the metal cage. I found these faeces contained *L. passim*. This indicates the parasites are excreted with the faeces and other bees may become infected by the ingestion. Similar horizontal transmission pathway was previously reported with *C. bombi* (Durrer and Schmid-Hempel, 1994). However, more data on the life cycle and transmission pathway of *L. passim* are necessary in the future.

2.3.2. Mortality of *A. mellifera* infected by *L. passim* under laboratory condition at different temperatures

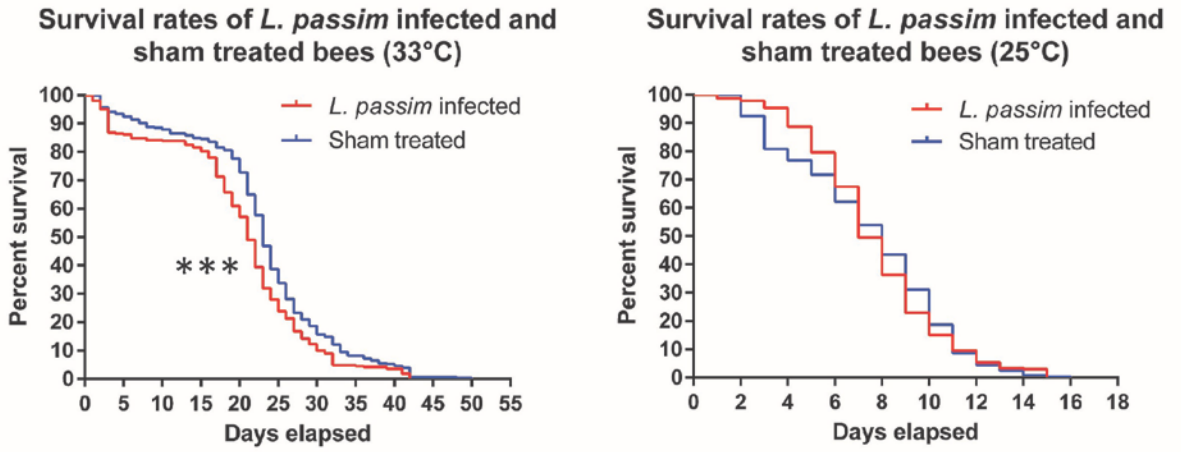


Figure 2.9 Mortality of honey bees infected by *L. passim* at 33 °C and 25 °C. At 33 °C, 300 infected and 300 sham treated bees were used for the test. At 25 °C, 240 infected and 240 sham treated bees were used for the test. Under each condition, three individual experiments were performed. The data were analyzed by Log-rank (Mantel-Cox) test. *P-value*=0.0002.

Since *L. passim* was not observed to penetrate the gut cells, the effect of *L. passim* infection on honey bee mortality was tested three times each at two different temperatures (Figure 2.9). At 33 °C under the laboratory condition, it took 42 days until all of the infected bees were dead. For the sham treated group, the bees survived up to 50 days. In the infected group, the median survival day is 21 days compared with 23 days for the sham treated group. Log-rank (Mantel-Cox) test shows that there is a significant difference between *L. passim* infected group and sham treated group. The survival rate of *L. passim* infected group is significantly lower than the sham treated group. From the individual mortality test, the mortality of *L. passim* infected bees was always higher than the sham treated bees when using Log-rank (Mantel-Cox) method. Thus, *L. passim* infection may slightly increase the mortality when fed with only sucrose. (See the individual mortality test at 33 °C in the Appendix 2). At 25 °C, there is no significant difference of mortality between *L. passim* infected and the sham treated groups according to the statistical analysis. Specifically, *L. passim* infected bees can survive up to 15 days. Similarly, the sham treated bees can survive up to 16 days. The median survival days were 7 and 8 days for the infected group and the sham treated group, respectively. At 25 °C, bees always form a cluster in the metal cage against the low temperature. Regardless of the considerable changes in ambient temperature, the temperature inside a hive is usually stable between 32 °C and 36 °C

(Williams et al., 2013). Thus, to maintain the bees at 33 °C should be ideal to simulate the natural condition for the further mortality test.

2.3.3. PCR detection of *L. passim* in the guts of honey bees tested for the mortality

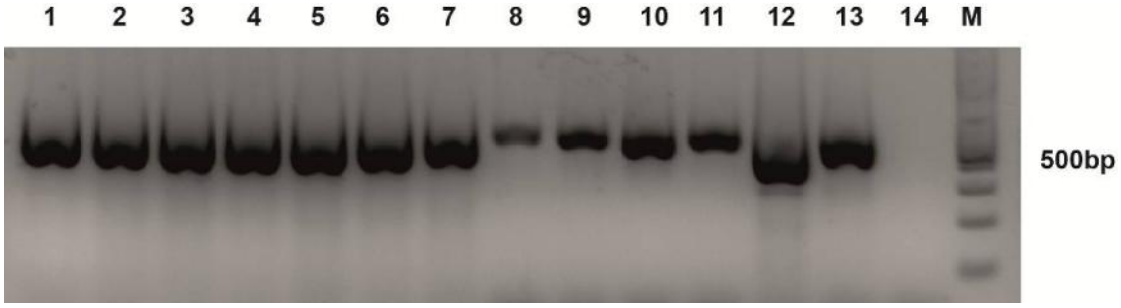


Figure 2.10 Results of PCR to detect the internal transcribed spacer 1 (ITS1, rRNA) of *L. passim* with seven genomic DNA samples extracted from the honey bee guts from the mortality test at 33 °C. Honey bee *AmHsTRPA* gene was used as a control to verify the quality of extracted DNA. Lanes 1-7 are the control PCRs to show *AmHsTRPA*. Lanes 8-14 are the ITS1 PCR to detect *L. passim*. Lanes 1-3 and 8-10 are from 15 days PI, lanes 4-7 and 11-14 are from 16 days PI. The lanes 7 and 14 are the results of PCR with the uninfected control sample and lane M is a molecular weight marker. As expected, the uninfected gut sample has no amplicon. However, all six infected gut samples have the amplicons with the expected size of 500 bp with different intensity, indicating the various degrees of infection in the individual honey bees.

In order to verify the infection of *L. passim* with the mortality test, six individual bees infected with *L. passim* together with the control bee were used for DNA extraction and PCR. Figure 2.10 shows that all of the infected bees have the amplicons around 500 bp. This indicates our method was effective to infect bees with *L. passim* under laboratory condition. But the infection levels are not constant.

2.3.4. qPCR analysis to measure *L. passim* abundance in the infected *A. mellifera* gut

As shown in Figure 2.11, results of measuring the abundance of *L. passim* in the honey bee gut after the infection demonstrate the number of parasites increased dramatically from day 1 to days 15 and 22. On days 15 and 22, the large error bars indicate there are large variations in the number of parasites between the infected individual bees. This result also indicates that *L. passim* starts to actively proliferate in the gut between day 8 and 15.

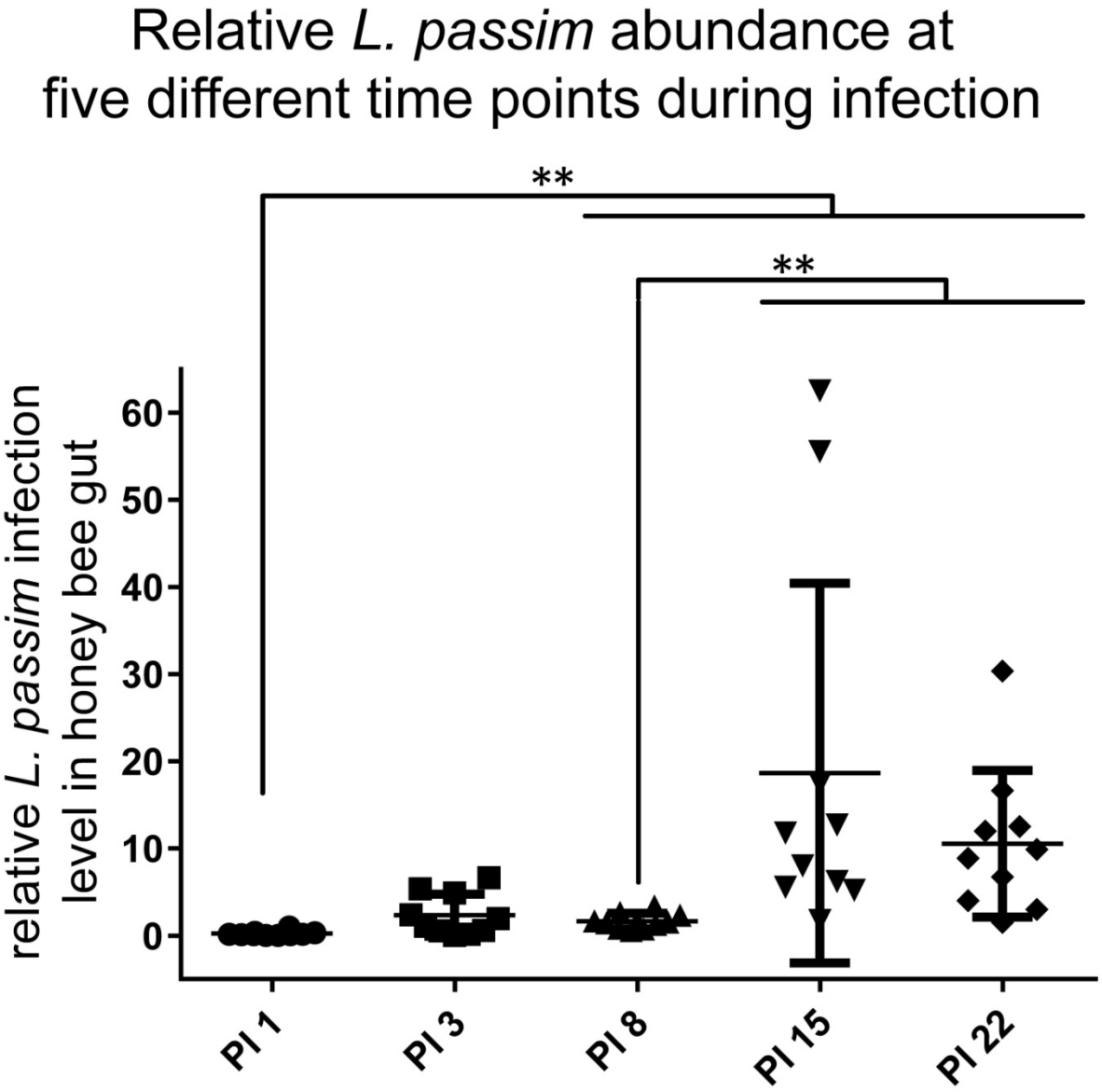


Figure 2.11 qPCR analysis to measure the abundance of *L. passim* in the ten individual infected bees at 3, 8, 15 and 22 days after the infection. In order to measure the relative abundance of *L. passim*, one honey bee sample at day 1 was set as 1. Only 9 samples were thus analyzed at day 1. Mean values±SD (error bars) are shown. ** P<0.01. For the statistical analysis, Steel-Dwass method was applied to compare different time points.

2.3.5. The effects of *L. passim* infection on *Firmicutes* and universal bacteria and the effects of pollen on *L. passim* load in honey bee gut

Since *L. passim* only colonized in the hindgut which also harbors various bacteria, different experiments were conducted in order to elucidate the potential effects of *L. passim* infection on the amount of both universal bacteria and *Firmicutes* since they were widely existed in the honey bee guts. Figure 2.12 shows that *L. passim* infection did not significantly affect the relative *Firmicutes* abundance in the infected honey bee guts compared with the control bee

guts. (Unpaired t-test, for *Firmicutes*: Day 7 PI: t = 1.130, P = 0.2773; Day 15 PI: t = 0.4364, P = 0.6692; Day 22 PI: t = 1.931, P = 0.0755; Day 27 PI: t = 0.1540, P = 0.8798). Similarly, the relative abundance of universal bacteria was not significantly affected by *L. passim* infection (Unpaired t-test, for universal bacteria: Day 7 PI: t = 0.8202, P = 0.4258; Day 15 PI: t = 1.635, P = 0.1243; Day 22 PI: t = 1.340, P = 0.2032; Day 27 PI: t = 1.193, P = 0.2526). However, pollen could protect honey bee from *L. passim* infection only in the early stage of infection (Unpaired t-test, for *L. passim*: Day 7 PI: t = 2.558, P = 0.0228; Day 15 PI: t = 2.675, P = 0.0181). Honey bees fed with pollen have significantly lower *L. passim* loads compared with the ones without pollen on day 7 and 15 under lab condition. In the late stage of infection, there is no difference between the group B and group C.; (Day 22 PI: t = 1.582, P = 0.1376; Day 27 PI: t = 0.5390, P = 0.5983).

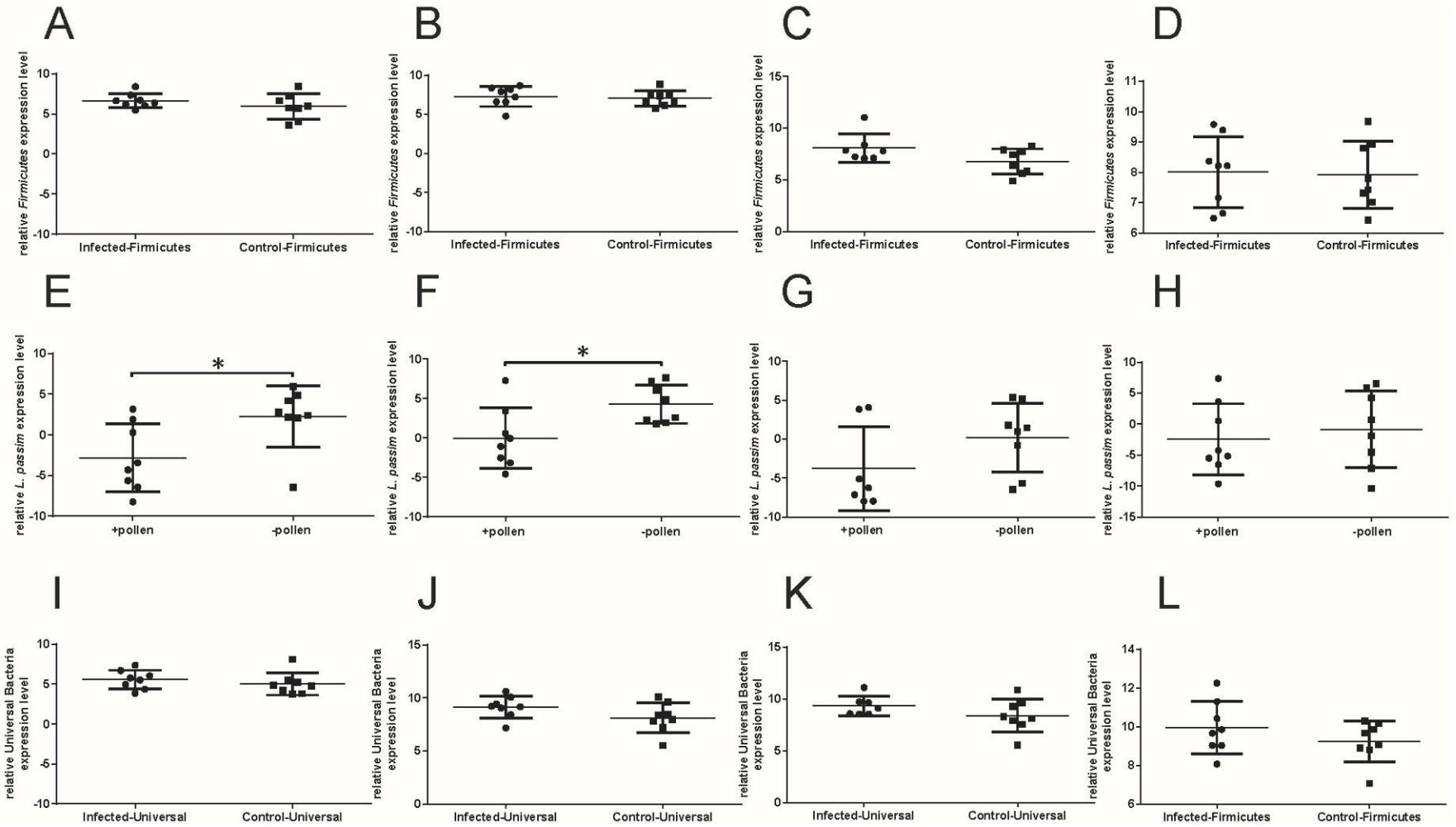


Figure 2.12 qPCR analysis of relative abundance for *Firmicutes*, universal bacteria and *L. passim* loads with/without pollen. A, B, C, D: Relative *Firmicutes* abundance at four different time points from *L. passim* infected and control groups. E, F, G, H: Relative *L. passim* abundance at four different time points with/without pollen. I, J, K, L: Relative abundance of universal bacteria at four different time points from *L. passim* infected and control groups. Statistical analysis was conducted by unpaired t-test with two-tailed calculation using GraphPad Prism 6.05. The asterisks represent two groups with a statistically significant difference ($p<0.05$). Mean values±SD (error bars) are shown. A, E, I: 7 days PI. B, F, J: 15 days PI. C, G, K: 22 days PI. D, H, L: 27 days PI.

The RNA-seq results described in Chapter 3 show that several nutrition and starvation-related genes were significantly up-regulated after 20 days PI (e.g. *take-out like carrier protein*, *facilitated trehalose transporter Tret1-like*). Enhanced expression of these mRNAs in the rectum of *L. passim*-infected honey bees led me to test *Vg* mRNA levels in the fat bodies. In Figure 2.13, the effects of *L. passim* infection on honey bee *Vg* mRNA expression indicates that the relative *Vg* mRNA expression in fat body was significantly lower in the infected than the control group under natural condition ($p<0.05$). All the samples were from two individual experiments. These results suggest that *L. passim* infected-honey bees are in poor nutritional status.

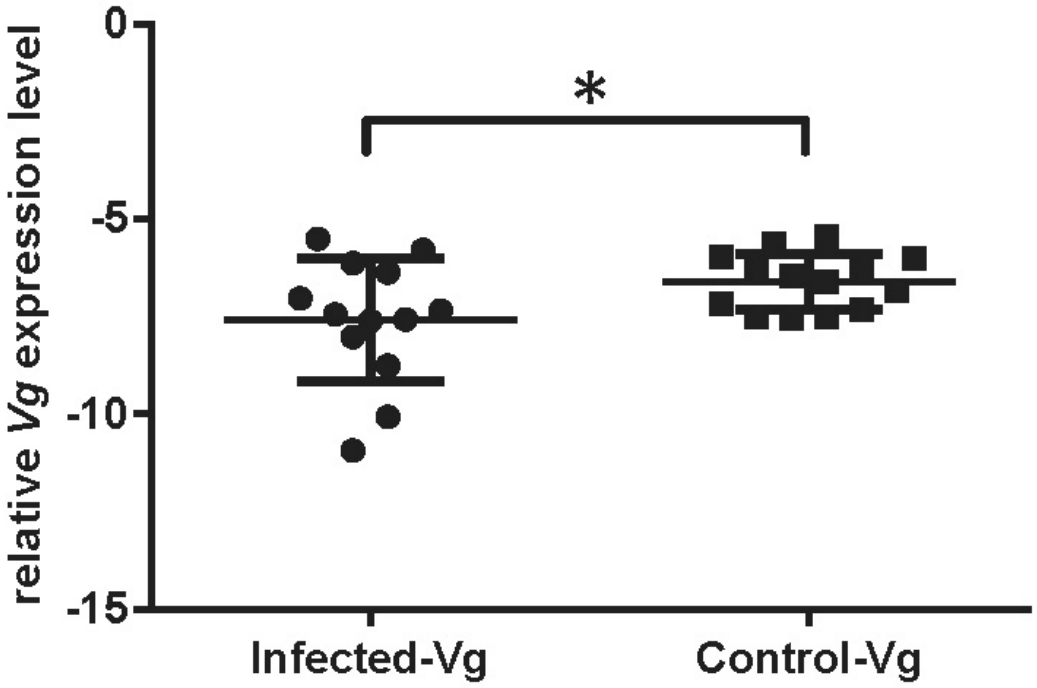


Figure 2.13 Relative *Vg* mRNA expression in the fat body of the infected and control honey bee at 20 days after the infection under natural condition. In total, 13 infected samples and 14 control samples were analyzed. All these samples were from two independent experiments. Statistical analysis was conducted by unpaired t-test with two-tailed calculation using GraphPad Prism 6.05. Asterisk represents two groups are statistically different ($p<0.05$). Mean values±SD (error bars) are shown.

2.3.6. The effects of *L. passim* infection on honey bee weight under natural condition

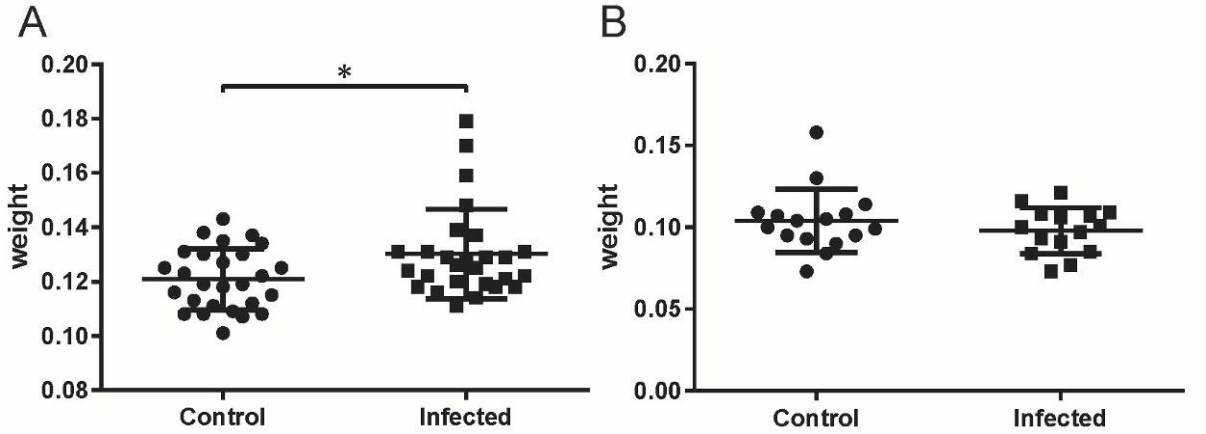


Figure 2.14 The effect of *L. passim* infection on honey bee weight at 20 days after the infection (A) and (B) indicate two individual experiments and each experiment was performed duplicate. Statistical analysis was conducted by unpaired t-test with two-tailed calculation using GraphPad Prism 6.05. The asterisks represent two groups with a statistically significant difference ($p<0.05$). Mean values±SD (error bars) are shown.

Figure 2.14 indicates the effect of *L. passim* infection on honey bee weight under natural condition. In the first experiment, the weight of *L. passim*-infected bees is significant higher than that of sham treated group (Unpaired t-test, t = 2.411, P = 0.0195). However, in the second experiment, there is no significant difference between the groups (Unpaired t-test, t = 1.005, P = 0.3232). These two individual experiments were performed in May and July 2018.

2.4. Discussion

Infection is initiated when honey bee ingests *L. passim* promastigotes which may come from contaminated flowers or nest mates by faecal-oral transmission pathway. *L. passim* passes through the honey bee digestive tract from midgut to hindgut. This procedure happens as short as three days. From my RNA-seq result, *L. passim* protein synthesis was decreased during infection which may help them to escape the immune surveillance of honey bee. From my RNA-seq results in the 3rd chapter, during the entire infection stage (day 7 to 27 PI), *L. passim* up-regulated several gp63 genes which play critical roles in establishing and maintaining infection. These genes can be used as acquiring nutrients from the honey bee guts and degrading the AMPs which released from honey bee. Also, *L. passim* has to migrate to the hindgut walls in order to attach to the cuticle cells. The *Lpgp63* gene may also play an important role in gut attachment. After attaching to the hindgut cuticle cells, *L. passim*

proliferate in the gut which is confirmed with my qPCR result (Figure 2.11). This step can generate more parasites which can be excreted through faeces. This is a transmission pathway of *L. passim* since the swimming parasites were found in the faeces from my experiment.

2.4.1. The impact of *L. passim* infection on honey bee mortality

L. passim infection slightly increase the mortality of bees compared with the sham treated group under the laboratory condition. If *L. passim* infection rapidly increases the mortality, it becomes difficult for *L. passim* to transmit to other bees. The low mortality of honey bee by *L. passim* infection would be critical for the high prevalence. Interestingly, *C. mellificae* infection did not significantly increase the mortality of infected honey bee compared with the control group either (Higes et al., 2016). This result is consistent with the one reported in the first paper about *C. mellificae* (Langridge and McGhee, 1967). In my experiment, *L. passim* infected honey bee can survive for 37 days under a natural condition during summer. However, during the over-wintering period, honey bee usually has a longer lifespan compared to summer. The effects of *L. passim* infection on honey bee during winter might be worth investigating since *L. passim* infection was reported as a negative factor for honey bee colony during overwintering period (Ravoet et al., 2013). Interestingly, previous report shows that pollen can significantly increase the *Vg* expression compared to the bees fed without pollen (Alaux et al., 2011). In my mortality test, only sucrose was provided. *L. passim* infection significantly decrease the expression of honey bee *Vg* which may reduce the longevity. Since it has been reported the expression of Vg can increase the lifespan of honey bee workers and queen (Corona et al., 2007;Nelson et al., 2007), *L. passim* infection can slightly increase the mortality may result from the suppression of honey bee Vg expression especially without pollen. This may indicate *L. passim* infection can slightly increase the mortality especially when the diet is limited (such as no pollen in the daily diet). However, monitoring the mortality of honey bee under a natural condition is challenging. At last, how does co-infection of *L. passim* with other pathogens/parasites affect honey bee health remains to be answered in future.

2.4.2. Impact of *L. passim* infection on honey bee gut bacterial community and the impact of pollen on *L. passim* infection

As described in the Results section, *L. passim* infection did not significantly alter the relative abundance of universal bacteria and *Firmicutes* on day 7, day 15, day 22 and day 27. However, it is still possible that the specific bacterial strains could be altered by *L. passim* infection. In this case, the species-specific primers should be tested in future. In my experiments, all bees were kept on the frames for 48 h before they were infected by *L. passim*. This indicates that the gut core bacteria were derived from the frame. My study demonstrates that *L. passim* does not modify the host fitness by altering the gut bacterial community in general. The recent experiment indicates that using *L. passim* treating honey bees induces a significant increase of total bacteria under hive condition (Schwarz et al., 2016). As the author proposed, the increased bacteria may result from a symbiotic relationship. However, under stressed condition (no protein, isolated from the hive), there is no significant difference between *L. passim* infection group and control group (Schwarz et al., 2016). The draft genome of *L. passim* indicates that *L. passim* has some unique genes which related to the carbohydrate metabolism. These genes also show bacterial origins which may be regarded as natural selection since *L. passim* can adapt to the honey bee gut environment which contains a lot of carbohydrates (Runckel et al., 2014). Thus, there are many factors, such as diet, social condition which can contribute to the honey bee gut bacteria. My results here did not show a significant difference for either universal bacteria or *Firmicutes* at four different time points under lab condition.

As mentioned in the Introduction, most bacteria species in honey bee gut can be cultivated *in vitro* (Kwong and Moran, 2016). This may provide another opportunity to feed newly emerged honey bees with specific bacteria species and subsequently infect with *L. passim*. Moreover, the different combinations of bacterial species can also be tested in order to precisely understand the interaction of honey bee gut bacteria with *L. passim*. In my experiment, all tested samples were honey bee workers. However, the effects of *L. passim* on honey bee queen are still unknown. Further direction would be to test if *L. passim* infection could alter the gut bacterial composition in the queen which may affect the fitness. Since the bacteria composition also shifts according to the age (> 30 days), there is no data from my experiments to show if *L. passim* could change the bacterial composition in the

elder workers. Furthermore, previous reports also imply that age, seasonal factor and different genetic background may also alter the ratio of the core gut bacteria species (Hroncova et al., 2015;Ludvigsen et al., 2015;Schwarz et al., 2016).

Recently, it has been shown that sunflower pollen can significantly reduce the *C. bombi* infection level and pollen also shows a long-term protective effect against *C. bombi* infection (Giacomini et al., 2018). The specific mechanism of how pollen can reduce the *C. bombi* infection level is still unknown. However, several potential mechanisms have been proposed. Firstly, the concentrations of fatty acids rich in the sunflower pollen might inhibit the growth of *C. bombi*. Secondly, the sunflower pollen may have the laxative properties which may decrease the time when parasites pass the gut. This may reduce the *C. bombi* attached to the hindgut thus reduce the infection level. The previous research indicated that the nectar alkaloids anabasine and nicotine can reduce the *C. bombi* infection which can be regarded as evidence to support this hypothesis (Richardson et al., 2015). Thirdly, recent *in vitro* data shows that the sunflower pollen extracts facilitate the growth of *C. bombi* rather than suppress the growth of *C. bombi* (Giacomini et al., 2018). This may indicate that the sunflower pollen may interact with the host which shows a protective effect *in vivo*. Most interestingly, since the sunflower belongs to Asteraceae, which is known to have spines on the outer coat (Blackmore et al., 2007). It has been proposed this pollen structure may help bumble bee to scrub the *C. bombi* which decrease the parasite load (Giacomini et al., 2018).

The pollen I used in the experiment was from a mixture of pollens with at least two distinct pollen species. This mixture of pollen shows a protective effect against *L. passim* in the early stage of infection. Figure 2.12 shows this protective effect lasts for at least 15 days compared with the experiments of sunflower pollen (7 days PI) (Giacomini et al., 2018). However, in the later stage of infection, the protective effect of pollen is absent. Although the pollen fed to the bees was not fresh, it can still decrease the relative abundance of *L. passim*. The pollen I used for the experiments also contains some with spine at the outer coats, indicating that these could be the reason for the protective effect from *L. passim* infection. In future, pure Asteraceae pollen species can be applied to decrease the *L. passim* infection level longer than 15 days. The previous report indicated the side effect of aged pollen to increase the mortality and affect the fitness of honey bees (Carroll et al., 2017). Fresh pollen can be used in future to test the effects on *L. passim* abundance. Different pollen species have different protein contents and also play an important role in honey bee health (Di

Pasquale et al., 2013;Palmer-Young et al., 2017). Bees rely on pollen for nutrients other than carbohydrates and phytochemicals. The previous study indicated that phytochemicals treatment can increase the expression of the antimicrobial peptide genes, such as hymenoptaecin (Palmer-Young et al., 2017). Nevertheless, my results indicate the pollen could be applied to decrease the *L. passim* infection level for at least a short-term.

The previous reports in *Drosophila* indicated that overexpression of AMPs results in an altered population of commensal bacterial communities in the gut. Specifically, some pathogenic bacterium was induced compared to the normal condition. The overexpression of this specific pathogenic bacterium ultimately leads to the apoptosis of gut cells and mortality (Ryu et al., 2008). In contrast, another report illustrated the expression of some AMPs is sufficient to rescue the resistance of immunodeficient fruit fly to infection (Tzou et al., 2002). Furthermore, the latest research shows that the overexpression of *Drosophila* AMPs (such as *Drosocin* gene) can extend the lifespan of the host by enhancing the gut immunity. Particularly, the extended lifespan was resulted from eliminating some opportunistic bacteria species as the author proposed. Unexpectedly, overexpression of *Drosocin* did not show any effects on the commensal microbiota in *Drosophila* midgut (Loch et al., 2017). From my RNA-seq results (see Chapter 3), it can be seen that *L. passim* infection triggers the expression of honey bee antimicrobial peptides (AMPs). *L. passim* infected bees did not show a shortened lifespan which may indicate these AMPs did not increase the mortality by impairing the gut homeostasis. However, how these AMPs affect the honey bee gut microbes should be investigated in future.

2.4.3. Effects of *L. passim* infection on honey bee weight

In the first experiment of weight measurement, *L. passim* infected group weighs heavier than the sham treated group under hive condition. However, in the second experiments, there is no significant difference between the *L. passim* infected group and sham treated group. It should be pointed out that these bees were from different hives and thus different genetic backgrounds might be the explanation for this inconsistency. In addition, this inconsistency may result from the outside temperatures. The outside temperature in July was higher than that of May. Due to the hot weather, bees are not willing to intake which might be a factor of this inconsistency. Previous report showed that *C. bombi* infection reduces the weight of the infected queens. *L. passim* infection may also reduce the honey

bee queen weight. However, more experiments should be tested in the future.

2.4.4. Honey bee *Vg* expression in fat body was reduced by *L. passim* infection

Under the hive condition, *L. passim* infection suppresses the expression of *Vg* mRNA compared with the control group. Honey bees infected with *L. passim* under hive condition have lower *Vg* expression and this may indicate they are more susceptible to the oxidative stress (Nelson et al., 2007). Suppression of *Vg* was also regarded as a symptom of senescence (Corona et al., 2007). Also, we may assume that *L. passim* infection may result in an early foraging behaviour under hive condition. This may be regarded as a protective strategy of social insect to minimize the transmission of other nest mates. The primary site of honey bee lipid metabolism (such as fatty acid synthesis and triacylglyceride) happens in fat bodies (Hahn and Denlinger, 2011). The suppression of *Vg* expression in the fat body may have negative effects on bee lipid metabolism. Moreover, many studies indicate that the Vg expression was considerably suppressed when lack of protein intake (Bitondi and Simoes, 1996;Cremonz et al., 1998;Alaux et al., 2011;Ament et al., 2011;Di Pasquale et al., 2013). Thus, *L. passim* infection could alter the honey bee nutritional state such as protein metabolism by decreasing the Vg in the fat bodies. Meanwhile, the lipid metabolism might be altered by *L. passim* infection. In the future, different concentration of *L. passim* can be inoculated to the bees to test the correlation between *Vg* expression as well as behaviours and the parasite loads.

3. Chapter III

Transcriptome analysis of *L. passim* and honey bee

3.1. Introduction of invertebrates response to pathogen infection

Insect innate immunity constitutes different levels of resistance through evolutionarily conserved defence strategies: physical barriers as the first line of defence, cell-mediated immunity and cell-free humoral immunity. These immunity responses are complex networks related to different signalling pathways which active various humoral factors. Among them, AMPs are recognized as essential components of humoral immunity among insects. The primary functions of AMPs are: 1) generate leaks in the membrane of pathogen/parasite; and 2) inhibit either folding or translation of proteins in pathogen/parasite. In honey bee, there are four families of AMPs, which are *apidaecins*, *abaecin*, *hymenoptaecin* and *defensins*. Specifically, as for honey bee *defensin*, it can sufficiently eliminate bacteria, fungi, microsporidium and viruses as reviewed before (Ilyasov et al., 2013). Moreover, a recent study also indicates the function of *defensin* related to the promotion of re-epithelisation not only in needle-wounded rats but also in adult human skin cells (Bucekova et al., 2017). In bumble bee, the increased expression of AMPs is induced by infection of Gram-negative bacteria (Erler et al., 2011). This may also indicate that AMPs might be involved in response to injury. In honey bee, there are several signalling pathways to control the expression of AMPs, i.e. Toll, IMD-JNK and JAK/STAT (Evans et al., 2006). By using RNA interference (RNAi) to silence *Relish*, which is a nuclear factor-B, researchers were able to show that *hymenoptaecin* expression was under the control of the Imd pathway. However, the expression of defensin-1 was not affected by *Relish* silencing, indicating the expression of *defensin-1* was regulated by Toll pathway instead (Schlüns and Crozier, 2007). Also, it should be pointed out that some AMPs could be regulated synergistically by both Toll and IMD pathways (Tanji et al., 2007). In the previous study with honey bee, the *NF-kappa-B inhibitor cactus 1* was also up-regulated when exposed to viral infection (Galbraith et al., 2015). In honey bee, *NF-kappa-B inhibitor cactus 1* gene was involved in both Imd and Toll signalling pathways which all related to the expression of AMPs (Brutscher et al., 2015). Thus, the up-regulation of *NF-kappa-B inhibitor cactus 1* may indicate the Imd and Toll signalling pathways should be inhibited which is regarded as an effect of pathogen invasion.

Bumble bee was shown to express several immune-related genes by *C. bombi* infection using RT-qPCR. These genes are *MyD88*, *Relish* and Thioester-containing protein 7 (TEP7) (Schlüns

et al., 2010). These genes were also reported to participate in the different signalling pathways of honey bees against viral infection (Brutscher et al., 2015). Similarly, the previous study reported the relative expression of several immune-related genes when honey bee was infected by *L. passim*. The relative gene expression of *L. passim* infection was examined from the whole abdomen and from the gut only according to the infection time. Specifically, in the early stage of *L. passim* infection (6-72 h), *L. passim* significantly induces Toll and Imd pathways by altering the expression of two signalling pathway components (*Imd* and *MyD88*) at the systemic level. Moreover, two genes (*Dscam* and *nimC1*) related to cellular recognition were also up-regulated by *L. passim* infection in the early stage (Schwarz and Evans, 2013). Furthermore, *L. passim* infection triggers the expression of AMP effectors genes. In the early stage of infection, *Defensin 2* was up-regulated systemically. However, 3 days after infection, the expression of *Abaecin* and *Defensin 1* was also triggered by *L. passim* infection. As for the tissue-specific gene expression, in the rectum, where *L. passim* was colonized, the relative expression of *Defensin 1* was significantly up-regulated compared with the uninfected control rectum (Schwarz and Evans, 2013). Nevertheless, RNA-seq was not used in this experiment, indicating honey bee genes induced by *L. passim* infection have not been thoroughly identified. Only the immune-related genes were analyzed. Thus, in this part of my project, RNA-seq was performed in order to gain a deep insight into the host (honey bee)-parasites (*L. passim*) interaction.

3.2. Materials and Methods

3.2.1. Biological samples and RNA preparations

New emerged honey bee workers either infected or sham treated were marked with different colours on the thorax using oil markers as described in 2.2.3. Subsequently, bees were placed back to the same hive to prevent the potential infection to the other hive which might affect the further infection experiments. At different time points, live bees from either infected or sham treated groups were taken back from the hive for RNA extraction. Specifically, live bees were immediately sacrificed by forceps and the whole abdomens were immersed in PBS. Then, the intact rectum was isolated and immediately homogenised by using chilled TRIzol® (Thermo Fisher). Generally, 1 µL total RNA was reverse transcribed using random hexamer primers with RTace reverse transcription enzyme from Takara as

described above. The complementary DNA (cDNA) library was used as templates to detect the *L. passim* infection. Specifically, *L. passim* glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) gene was used to detect the infection from the total RNA (Kojima et al., 2011). Honey bee elongation factor 1 (*Ef-1alpha*) was used as a positive control to verify the reverse transcription (Kojima et al., 2011). 10mL medium with free swimming parasites was centrifuged at 2,000 x g for 5 min at room temperature. The supernatant was discarded before 1 mL of TRIzol was added. Under the natural conditions, 2 infected and 2 sham treated honey bee from 7 days PI, 12 days PI, 20 days PI and 27 days PI were used for total RNA extraction. Figure 3.1 shows how the RNA-seq experiment was designed. Intact rectums were used for total RNA extraction for RNA-seq. All the infected samples were quantified by qPCR or PCR to guarantee the relative same infection level of *L. passim* in the honey bee guts. Eighteen samples above were sent to Beijing Genomics Institute (BGI) Wuhan using Illumina HiSeq-4000 platform for sequencing. Total RNA samples from control bee guts were mixed if the quantity was not enough for normal RNA-seq. All total RNA samples met the highest standards on both quality and quantity to ensure the accuracy of the outputs. For each sample, at least 4 GB clean reads were obtained for the downstream analysis. The raw sequencing data was deposited to NCBI under the accession number PRJNA510495 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA510495).

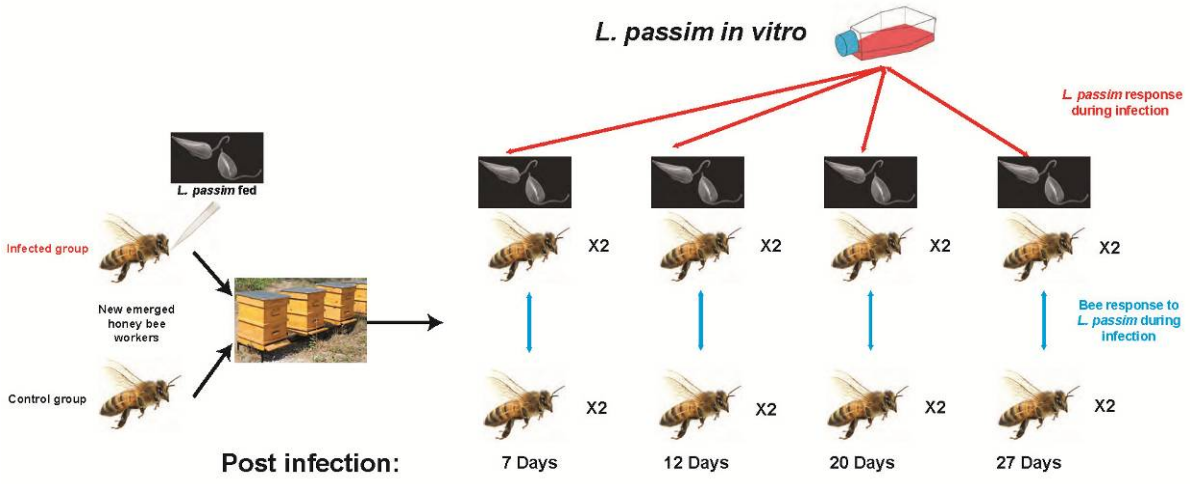


Figure 3.1 Outline of the experimental design to analyze *L. passim* responses during the infection in honey bee and the honey bee response to *L. passim* infection at different time points. Red lines indicate the comparisons of *L. passim* samples whereas blue lines indicate the comparisons of honey bee samples. There are two honey bee samples for both *L. passim* infected and control at each time point for RNA-seq.

3.2.2. RNA-seq data analysis for honey bee and *L. passim*

As for honey bee RNA-seq analysis, the reference genome assembly (GCF_000002195.4_Amel_4.5_genomic.fna), annotation GFF file (GCF_000002195.4_Amel_4.5_genomic.gff) and the protein fasta (GCF_000002195.4_Amel_4.5_protein.faa) files of honey bee were downloaded from National Center for Biotechnology Information (NCBI). The program gffread (http://ccb.jhu.edu/software/stringtie/gff.shtml) was used to convert the honey bee GFF file to GTF for HISAT2-build on a local MacBook Pro. The splice junctions and the positions of exons were extracted from the converted GTF file by using hisat2_extract_splice_sites.py and hisat2_extract_exons.py Python scripts which from HISAT2 (Version: 2.1.0) (Kim et al., 2015). Next, HISAT2-build was used to build the HISAT2 index with the spliced sites and exons. Then, clean reads were mapped to the latest version of the honey bee index. Then, SAM files were converted and sorted by SAMtools (Li et al., 2009). HTSeq-count was used to count from each sorted SAM file. NCBI RefSeq name was used as a unique identifier to count the reads from SAM files. The genome annotation of honey bee was downloaded in tabular format which used to align the RefSeq Name identifier to the functional annotation. Similarly, the genomic sequence of *L. passim*, strain SF (GenBank number: AHIJ00000000.1) was downloaded from NCBI (Runckel et al., 2014) to build the HISAT2 index. The annotation GFF file which used for HTSeq-count was generated from Companion pipeline (Steinbiss et al., 2016) as described in 3.2.3. The protein fasta file from Companion pipeline was used to align the protein function of each gene.

3.2.3. Re-annotation of *L. passim* by Companion pipeline

Clean reads were mapped to the *L. passim* genome using TopHat V2.1.0 with default parameters (Trapnell et al., 2009). All transcripts containing *L. passim* were assembled using Cufflinks v2.2.1 (Trapnell et al., 2012). Once the short read sequences were assembled, all the transcript evidence were merged using Cuffmerge (Trapnell et al., 2012). The merged *L. passim* transcript evidence and the genomic sequence of *L. passim* were annotated by using the parasite genome annotation pipeline Companion with all *Trypanosomatida* species: *L. major Friedlin*, *L. braziliensis* and *T. brucei TREU927* as reference organisms (Steinbiss et al., 2016). In the pseudochromosome contiguation part, the minimum required match length for

contig placement was set to 200 bp and the minimum similarity for contig placement were set to 35%. In the advanced settings step, reference proteins were aligned to the target sequence. The outputs of the companion pipeline were visualized by Artemis to compare the unassembled "bin" chromosome size to determine the closest species of *L. passim* (Rutherford et al., 2000). After determining the best reference species for *L. passim*, the re-annotated GFF3 file and the protein coding sequences were used for the Differential Gene Expression (DGE) and other downstream analysis. Latest *L. passim* annotation gff3 file and genome assembly were uploaded to the online eukaryotic pathogen CRISPR guide RNA/DNA design tool (http://grna.ctegd.uga.edu/) in order to predict the potential off-target effects in *L. passim* CRISPR/Cas9 system (Peng et al., 2015). All the hypothetical proteins in *L. passim* from Companion pipeline was extracted and imported to Blast2Go again to run against the nr database with the same parameters as described above to retrieve more annotations. The BLAST results of these hypothetical proteins combined with the old annotation file to obtain the relative complete annotation of *L. passim*.

3.2.4. DGE analysis, sample quality assessment of *L. passim* and honey bee

Raw read counts from both *L. passim* and honey bee were analyzed in RStudio (Version 3.4.3) using the generalized linear model (GLM) based method of edgeR package (Version 3.20.9) from Bioconductor (Robinson et al., 2010). Normalization was performed with the Trimmed Mean of M-values (TMM) which implemented in the edgeR Bioconductor package (Robinson and Oshlack, 2010). A filter was set up to remove any read counts less than one count per million mapped reads (CPM). After testing, the Benjamini-Hochberg (BH) method was applied to control the false discovery rate (FDR) across the detected loci. As for *L. passim,* the raw *P-value*s for each gene were corrected for multiple setting an FDR of 0.01 which is considered to be differentially expressed. As for honey bee, the FDR was set up to 0.05 which were considered to be DEGs among different time points under different conditions. A multi-dimensional scaling (MDS) plot and Pearson correlation-based heat maps created using the gplots pheatmap.2 function which was used to calculate a 2D projection of distances between samples. After obtaining the DGE, Venny (http://bioinfogp.cnb.csic.es/tools/venny/index.html) was used to make the Venn diagrams

to summarize the proportion of the global DGE of *L. passim* and honey bee under different conditions and visualized them.

3.2.5. Gene ontology (GO) enrichment analysis (Fisher's Exact Test)

In order to minimize the computational time and increase the specificity of the analysis, local blast databases were made using Blast2Go software with the taxonomy IDs. In total, 446,413 protein sequences from *Order Kinetoplastida* (Taxonomy ID: 5653) and 6,225,195 protein sequences from *Subphylum Hexapoda* (Taxonomy ID: 6960) were downloaded from NCBI database (https://www.ncbi.nlm.nih.gov/) with GI number. All these sequences were concatenated into two FASTA files separately. Then, these two FASTA files were used to build the local BLAST database in Blast2GO PRO desktop version 5.0.1 (Conesa et al., 2005). After obtaining the DGE from *L. passim* and honey bee, up-regulated and down-regulated genes were analyzed individually at different time points. For each data set, protein sequences were extracted from the whole protein FASTA file of *L. passim* or *A. mellifera* using python script and imported to Blast2GO. Then, local BLAST was performed using program blastp (parameters: E-Value Hit Filter = 1e-10; Number of Blast Hits: 20) against proper local BLAST local database which created above. Simultaneously, for each dataset, InterPro was performed using the EMBL-EBI public web-service database with the default settings. Next, mapping and annotation procedures were carried out using the latest database online according to the default settings. After annotation, the InterProScan GOs were merged to annotation. Finally, enrichment analysis (Fisher's Exact Test) was performed using the program which embedded in the Blast2GO. The entire protein-coding sequences from *L. passim* and honey bee were used to generate the reference annotation lists for the GO enrichment analysis. Different lists from different DGE were used as test-set files for the GO enrichment analysis. After obtaining the GO enrichment terms, all the GO enrichment terms were reduced to most specific GO terms by setting 0.05 as FDR cut-off. The complete GO enrichment results can be found in the supplementary documents.

3.2.6. DEXseq analysis of different exon usage by *L. passim* infection

The genome assembly (Apis_mellifera.Amel_4.5.dna.toplevel), annotation GTF and GFF files (Apis_mellifera.Amel_4.5.38.gtf, Apis_mellifera.Amel_4.5.38.gff3) of *A. mellifera* from the latest release were downloaded from EnsemblMetazoa FTP website

(https://metazoa.ensembl.org/info/website/ftp/index.html). The SAM files from HISAT2 (Kim et al., 2015) were sorted by SAMtools (Li et al., 2009). The Python script dexseq_prepare_annotation.py in DEXseq package (Anders et al., 2012) was used to convert the Ensembl GTF file and translates it into a GFF file with collapsed exon counting bins based on HTSeq (Anders et al., 2015). For each SAM file, python_count.py (HTSeq) counts the number of reads that overlap with each of the exon counting bins defined in the flattened GFF file. Full model of counts = sample + exon + condition: exon was used to test the differential exon usage between *L. passim* infected bees and control bees by DEXseq package in R. This characterized genes which show differential exon usage due to the interaction of *L. passim* infection (FDR < 0.01). The non-coding RNA was removed from the analysis and the full gene list from DEXseq outputs was submitted to DAVID (Huang et al., 2008b;a) to retrieve the description of each gene. The genes with no records were marked with NA and kept in the list of DEXseq output. The full lists were sorted by exon_change which directly from DEXseq analysis.

3.3. Results and Discussion

3.3.1. Reference organism selection for re-annotation of *L. passim* genes

Comparing the macro synteny of *L. passim* and *T. brucei* genomes, only 4 pseudochromosomes of *L. passim* were placed on *T. brucei* genome. However, for *L. braziliensis* and *L. major*, macrosynteny of *L. passim* genome is shared with these of two *Leishmania* reference chromosomes. Specifically, 36 pseudochromosomes were placed on the genome of *L. passim*. In this case, compared with *T. brucei*, *L. passim* should be a close relative to *L. major* and *L. braziliensis*. Not surprisingly, *T. brucei* is always used as an outgroup of all Leishmaniiinae species (Schwarz et al., 2015). Importantly, for *L. major* and *L. braziliensis*, the unassembled "bin" chromosome sizes were 18,857,718bp and 21,308,849bp, respectively indicating more contigs were placed to *L. major* compared to *L. braziliensis*. In total, 36 pseudochromosomes were annotated by Companion pipeline. A total of 9342 protein-coding sequences (CDS) were predicted by Companion supported with the transcript evidence. Among these, 3,919 were annotated as "hypothetical proteins". After running blast against the nr database, only 330 of them were annotated with some functions. The

combined annotation file was used to annotate the genes from the edgeR outputs. The *L. passim* protein file was used for the downstream analysis. The *L. passim* annotation file (gff3) was used for all RNA-seq analysis.

3.3.2. *L. passim* RNA-seq statistics and the relationship between the samples

For the *L. passim*-infected honey bee samples, a total of 2.33×10^8 raw reads were generated after removing the low-quality reads by BGI. A total of 1.45×10^8 reads were aligned to the *L. passim* genome (61.37% alignment rate). In the edgeR analysis (Robinson et al., 2010), a total of 7.19×10^7 reads were used for DGE analysis after removing the reads with very low expression. For the control honey bee samples without *L. passim* infection, a total of 3.28×10^8 raw reads were generated after removing the low quality reads by BGI. A total of 1.53×10^8 reads were aligned to the *A. mellifera* genome (55.10% alignment rate). In the edgeR analysis (Robinson et al., 2010), a total of 1.07×10^8 reads were used for DGE analysis after removing the reads with low expression. All HISAT2 alignment statistics, the edgeR library sizes from both *L. passim* and honey bee samples can be found in the Appendix 3.

MDS plot and Pearson correlation heat maps were used to characterize the correlations between all RNA-seq samples, the reproducibility of duplicate, and the potential batch effects among all samples. Gene expression profiles of *L. passim* during the infection in honey bee can be classified into 4 clusters. Specifically, the duplicate of parasites cultured with the liquid medium was classified into the single separate cluster. The analysis also revealed a high degree of similarity between the duplicate samples regardless of the library size variations. Moreover, the MDS plot also indicates a low batch effect with no clustering based on biological replicate but rather on the infectious stages of *L. passim* in the honey bee gut. This indicates that the gene expression profiles of *L. passim* change when they are cultured with the medium or at the different stages, early (7 days PI), middle (12 days PI) and late stages (20+ days PI) of the infection in honey bee. This may also indicate the physiological states of *L. passim* change in order to adapt to the microenvironment of honey bee gut at different time points of the infection.

3.3.3. Identification of DEGs of *L. passim* during infection in the honey bee gut

For *L. passim*, a total of 3,240 genes were differentially expressed during infection in the honey bee guts compared to the parasites cultured in the medium. Among these, 1,529 genes were up-regulated whereas 1,711 genes were down-regulated at least once during the infection. Figure 3.2 shows the DEG of *L. passim* during infection in the honey bee gut. The Venn diagram shows that 44 genes (2.9%) and 67 genes (3.8%) of *L. passim* were up- and down-regulated, respectively throughout the entire infection period. These genes may play critical roles for *L. passim* to establish and maintain infection in the honey bee gut microenvironment. The number of up- and down-regulated genes was the highest at 12 days after the infection. This may indicate at this time point, *L. passim* dramatically changes the gene expression to establish the continuous infection in the honey bee guts. Furthermore, 262 genes (17.1% of all up-regulated genes) were up-regulated at the early stage of infection (7 days after the infection). These genes may critical for *L. passim* to initiate the infection in the gut and also to escape from the host immune responses. The complete lists of DEG of both *L. passim* and honey bee at four different time points can be found in the supplementary documents.

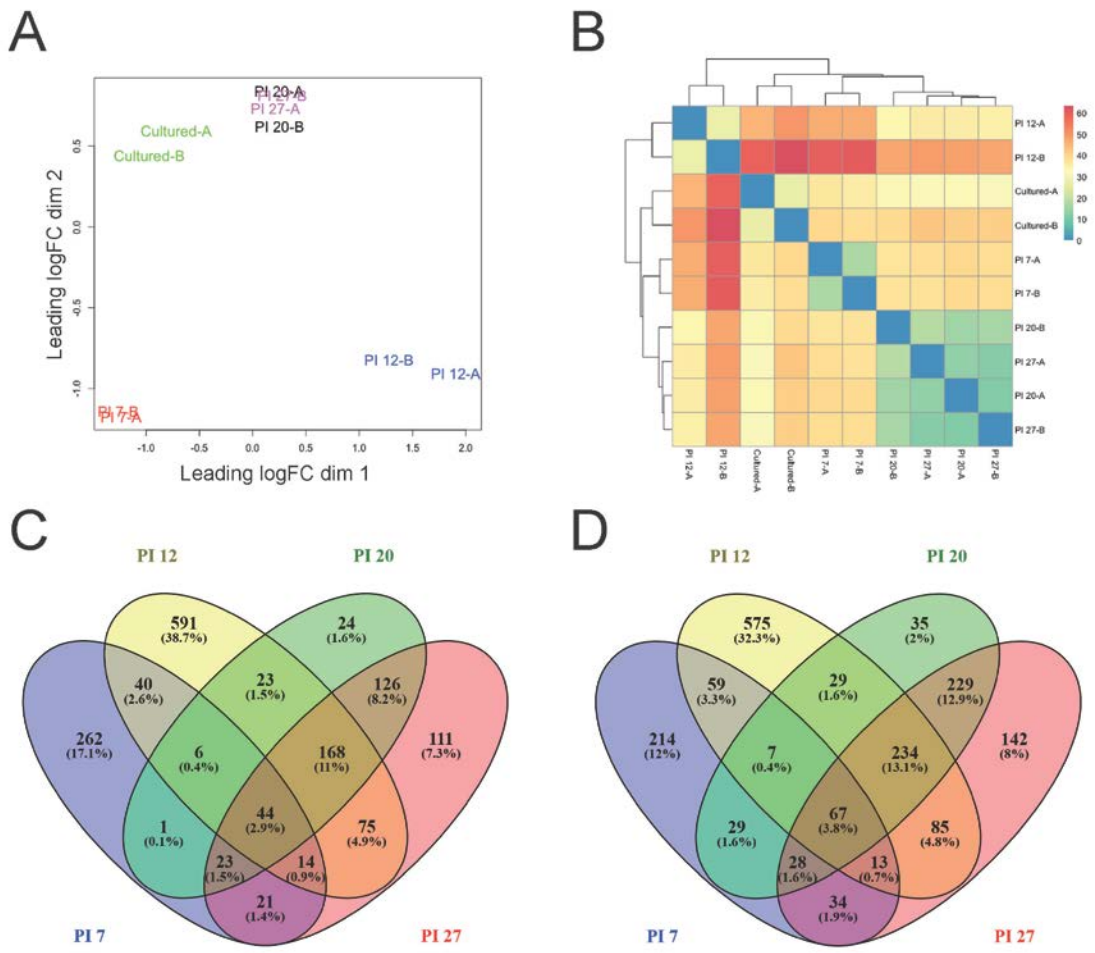


Figure 3.2 (A) Multi-dimensional scaling (MDS) plot to illustrate a 2D projection of differential gene expression between samples. Samples collected at different time points during infection (PI 7-27) as well as the parasites cultured with the medium are indicated by different colors. Two replicates are marked as A and B, respectively. (B) Sample hierarchical analysis heat map using Pearson correlation. The levels of difference between samples are indicated by different colors. (C) Venn diagram to indicate *L. passim* genes up-regulated at four different time points (PI 7-27) during infection ($P < 0.01$). (D) Venn diagram to indicate *L. passim* genes down-regulated at four different time points (PI 7-27) during infection ($P < 0.01$).

3.3.4. GO enrichment analysis of the DEGs of *L. passim* during the infection

GO enrichment analysis was performed to characterize the general functions associated with the DEGs of *L. passim* during the infection in honey bee gut. Figure 3.3 illustrates the GO enrichment terms from all up-regulated genes of *L. passim*. Seven GO terms were enriched with the up-regulated genes throughout the entire infection period. These GO terms include: proteolysis (GO:0006508), modulation by symbiont of host protein kinase-mediated signal transduction (GO:0075130), modulation by symbiont of host nitric oxide-mediated signal

transduction (GO:0044081), host cell nucleus (GO:0042025), metallopeptidase activity (GO:0008237), carboxypeptidase activity (GO:0004180) and protein tyrosine phosphatase activator activity (GO:0008160). Nitric oxide (NO) is a critical molecule which participates in different signalling pathways. NO is generated by the oxidation of L-arginine to citrulline catalyzed by the NO synthase (Rivero, 2006). In honey bees, many reports have shown that NO plays a critical role as a neurotransmitter in learning (Müller, 1996;Müller and Hildebrandt, 2002). More recently, NO was reported in the insect immunity when mosquito and fruit fly were challenged by bacteria in the gut (Davies and Dow, 2009;Hillyer and Estévez-Lao, 2010). In *Drosophila paramelanica*, when the larvae were parasitized by wasp, a dramatic increase of NO was observed indicating the participation of NO in the early stage in the response of infection. Furthermore, the author proposed that NO might play a role in recruiting hemocytes to the infection sites (Carton et al., 2009). In honey bee, it has been reported that NO plays an important role at the beginning of the immune response to identify non-self in the hemocytes. The generation of NO triggers the spreading of hemocyte which is believed as the first step of immune activation after recognition of foreign surfaces (Negri et al., 2013). As for the protein tyrosine phosphatase (PTP) activator activity, it has been reported that *Leishmania* gp63 activates host Src-homology 2 domain-containing phosphatase (SHP-1) and PTP1B which inhibits the host's immune pathways such as JAK2/STAT pathway. Also, the activation of SHP-1 and PTP1B also inhibits the production of NO (Shio and Olivier, 2010). In these two GO terms, two *L. passim* gp63 genes participated in the test set indicating critical roles of *Lpgp63* during infection.

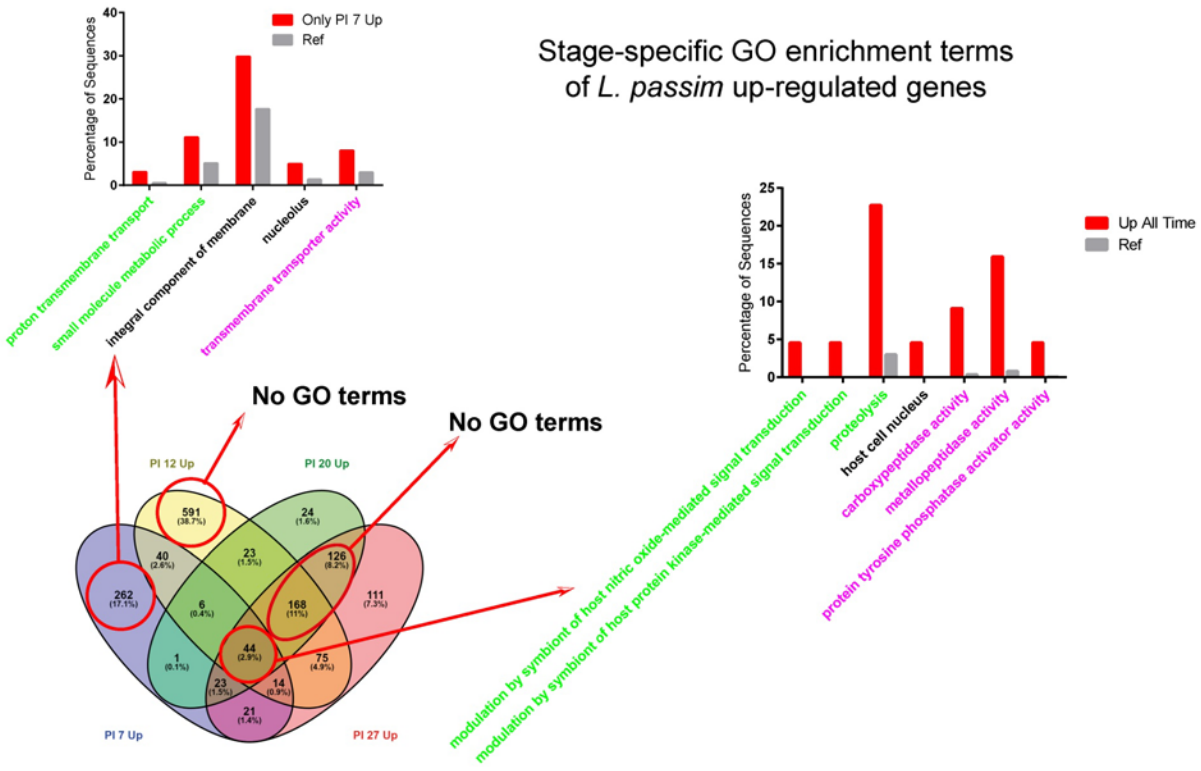


Figure 3.3. Stage-specific GO enrichment terms of *L. passim* up-regulated genes during the infection. The Venn diagram indicates all of the *L. passim* up-regulated genes during the infection. Four time points analyzed are 7 days PI, 12 days PI, 20 days PI and 27 days PI. The 262 and 591 up-regulated genes would be important at the early-mid stage of infection. The 126 and 168 up-regulated genes would be important at the mid-late stage of infection. In the middle of the Venn diagram, 44 genes which are up-regulated across the entire infection period. In each bar chart, the red bars indicate the test gene sets and the grey bars are the reference gene sets of *L. passim*. All of the GO terms are classified into 3 categories and highlighted with different colours on the X-axis which are green (Biological Process), black (Cellular Components) and magenta (Molecular Function). Y-axis represents the percentage of sequences.

Figure 3.4 indicates that only one GO term, cellular homeostasis (GO:0019725) (*P-value*: 8.37E-06), was enriched with the down-regulated genes throughout the entire infection period. These genes include: tryparedoxin peroxidase (Lp_000309900.1), thiol-dependent reductase 1 (Lp_000481500.1), tryparedoxin 1, putative (Lp_000162900.1), tryparedoxin peroxidase (Lp_000310000.1), tryparedoxin-like protein (Lp_290008900.1).

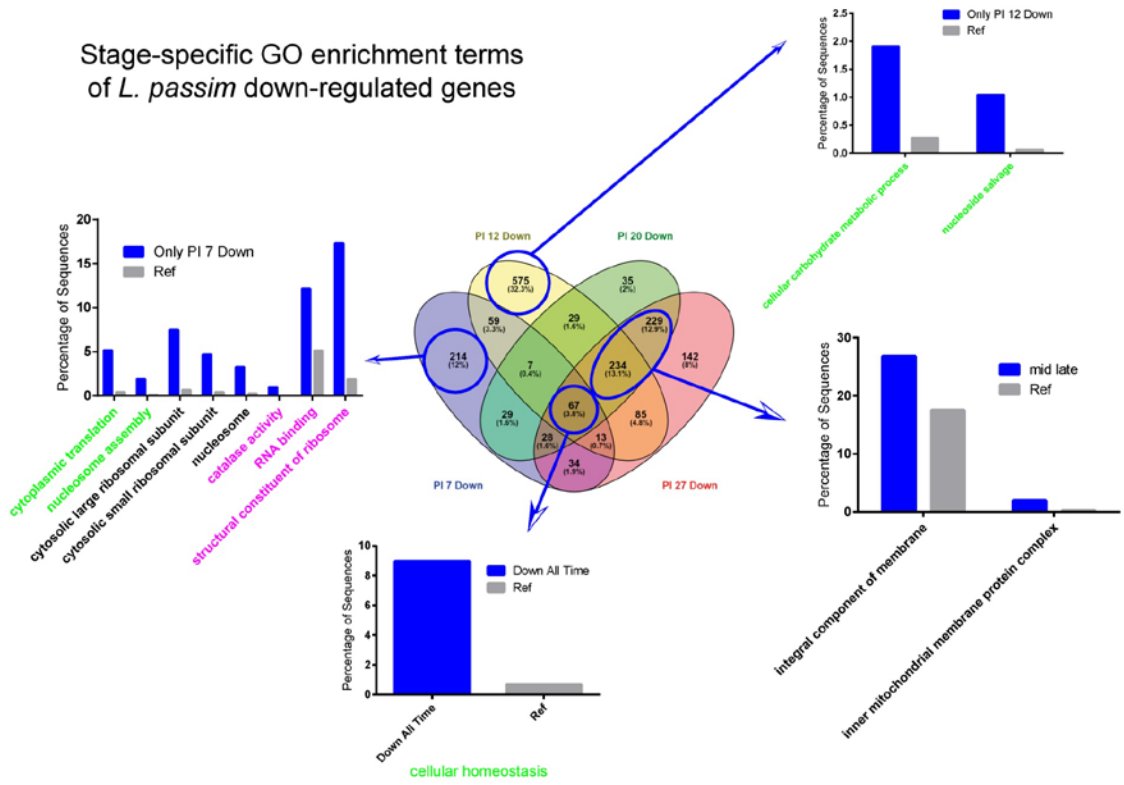


Figure 3.4. Stage-specific GO enrichment terms of *L. passim* down-regulated genes during the infection. The Venn diagram indicates all the *L. passim* down-regulated genes during infection. Four time points analyzed are 7 days PI, 12 days PI, 20 days PI and 27 days PI. The 214 and 575 down-regulated genes would be important at the early-mid stage of infection. The 234 and 229 down-regulated genes would be important at the mid-late stage of infection. In the middle of the Venn diagram, 67 genes which are down-regulated across the entire infection period. In each bar chart, the blue bars indicate the test gene sets and the grey bars are the reference gene sets of *L. passim*. All the GO terms are classified into 3 categories and highlighted in different colours on the X-axis which are green (Biological Process), black (Cellular Components) and magenta (Molecular Function). Y-axis represents the percentage of sequences.

In trypanosomatid, it has a unique trypanothione-based thiol metabolism which is different from any other species. Trypanothione is an unusual form of glutathione which uses spermidine to conjugate two glutathione together (Fairlamb and Cerami, 1992). The tryparedoxin/tryparedoxin peroxidase system (TXN/TXNPx) uses Trypanothione as a source of electrons. The previous paper indicates that TXN activates the UMSBP (oxidized form) substrate binding to UMS (Onn et al., 2004). It has been shown in *T. brucei* that overexpression of *TbTXNPx* results in the loss of kDNA which due to the increase in UMSBP oligomerization (Motyka et al., 2006). In *Crithidia fasciculata*, it has been shown that the TXNPx inhibits the UMSBP binding to the UMS DNA (Sela et al., 2008). This system is critical for kinetoplastida since they use a cascade of enzymes to detoxify the oxidative stresses during the development in different hosts and vectors (Fiorillo et al., 2012). Moreover, the recent studies also demonstrate the critical role of TXN/TXNPx system which controls the redox state of the transcription factor UMSBP (Sela et al., 2008;Castro et al., 2010;Shlomai,

2010). Take *L. donovani* as an example, the redox mitochondrial enzymes TXN and TXNPx regulate the UMSBP binding with UMS. Specifically, *L. donovani* TXN and TXNPx oppositely regulate the binding activity of UMSBP to UMS. UMSBP KO in *L. donovani* results in the loss of kDNA, then decrease the expression of cytochrome b. Since cytochrome b is an essential component of the electron transport chain (ETC), the down-regulation of cytochrome b leads to the disruption of complex-III activation. In the end, the UMSBP KO results in the decrease of ATP generation by oxidative phosphorylation which has negative effects on the parasite (Singh et al., 2016). In this case, it can be seen that UMSBP plays a central role in the regulation binding activity to UMS which ultimately affects the ATP generation. The proposed schematic representation of trypanothione metabolism and the cascade of enzymes involved in the utilization of trypanothione in *L. passim* are shown in figure 3.5. The down-regulation of *L. passim* tryparedoxin peroxidase (LpTXNPx) during infection also indicates the enhancement of *L. passim* universal minicircle sequence binding protein (LpUMSBP) binding activity to *L. passim* universal minicircle sequence (LpUMS). Since the honey bee hindgut is under an anaerobic condition, the amount of oxygen is limited. *L. passim* needs to up-regulate UMSBP binding to UMS in order to maximize the oxidative phosphorylation at the early stage of infection. Indeed, the down-regulation of *LpTXNPx* can be found throughout the infection. Since the primary function of UMSBP is related to the replication of kDNA, we could also assume the up-regulation of *LpUMSBP* may also cooperate with *L. passim* structure-specific endonuclease I (LpSSE-I) for initiating the replication of kDNA since this gene was also up-regulated. The up-regulation of *L. passim* UMSBP and SSE-1 stimulate kDNA replication and thus enhance the expression of cytochrome b, protein complex-III on ETC to stimulate ATP synthesis. Our results indicate the importance of *LpTXN*/*LpTXNPx* pathway and how these specific redox reactions interact with *LpUMSBP*.

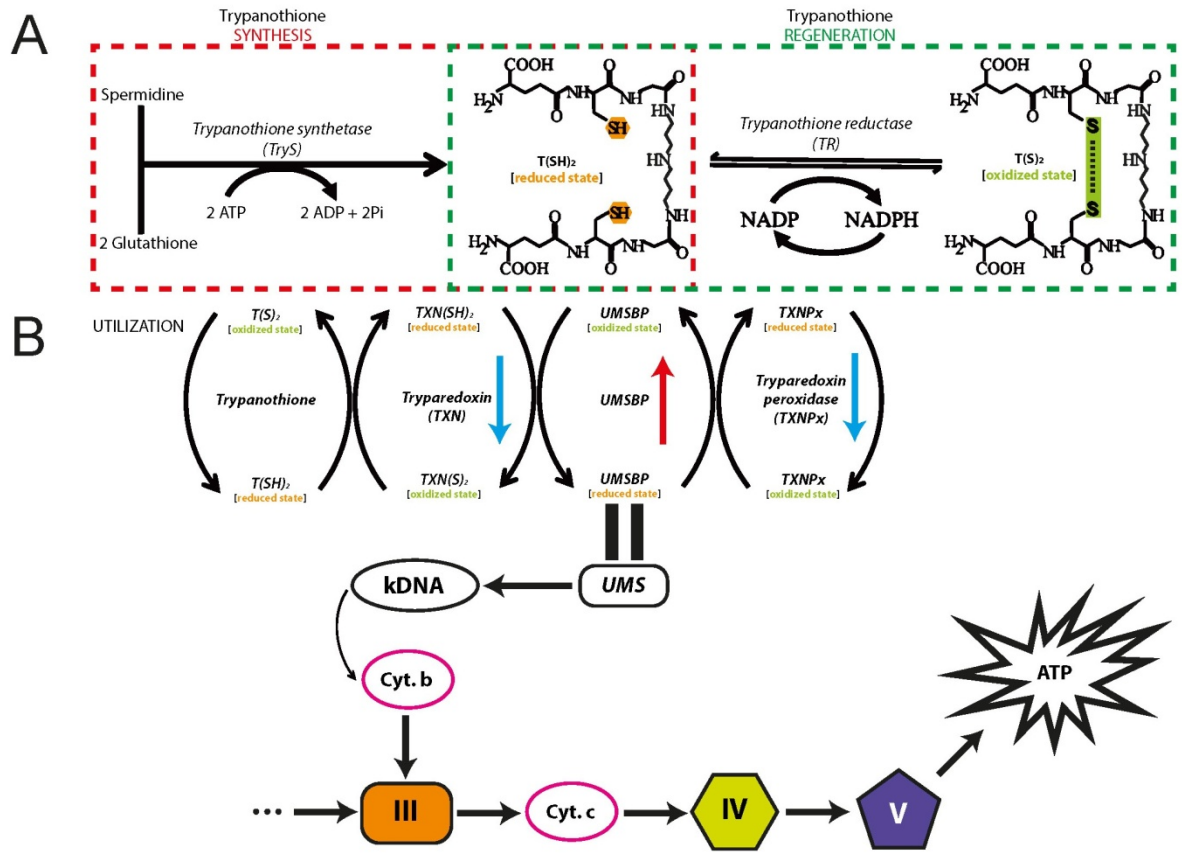


Figure 3.5 Unique trypanothione dependent redox metabolism and gene regulation during infection. (A) The synthesis and regeneration of trypanothione were shown. Synthesis: two molecules of glutathione and one molecule of spermidine are ligated together by trypanothione synthetase (TryS) at the expense of ATP. Regeneration: trypanothione reductase (TR) maintains trypanothione in the reduced state at the expense of NADPH. The synthesis and regeneration processes are shown in red and green dashed-line rectangles. (B) The usage of trypanothione in the cascade of enzymes and their roles in UMSBP, kDNA, ETC and ATP generation. The reduced form of trypanothione participates in the cascade of enzymes which related to the binding of UMSBP to UMS. The tryparedoxin (TXN) and tryparedoxin peroxidase (TXNPx) redox system regulate the binding of UMSBP to the replication origin UMS through the opposing effects. Reduced UMSBP binds with UMS of minicircles and the replication of kDNA initiated. The kDNA encodes the mitochondrial genes which have mitochondrial functions and these genes participate in the ETC. Cyt.b is a critical component of ETC Complex III and a cascade of enzymes ultimately lead to ATP generation. During *L. passim* infection, *LpTXN*, *LpTXNPx* down-regulated all across infection and the *LpUMSBP* up-regulated only on day 7 PI. All these differentially expressed genes indicate the enhancement of UMSBP binding activity and the copy number of kDNA might be increased to maximize the oxidative phosphorylation. The red and blue arrows indicate the up- and down-regulation of genes during infection. Adapted and modified from (Benítez et al., 2016;Singh et al., 2016).

All of the enriched GO terms in the MF category which retrieved from *L. passim* up-regulated genes throughout infection are all subgroups of hydrolase activity. Among them, the most specific ones are metallopeptidase activity (GO:0008237) (*P-value*: 7.03E-08) and carboxypeptidase activity (GO:0004180) (*P-value*: 2.02E-05). Aminopeptidases, carboxypeptidases and metallopeptidases were involved in the process of protein degradation. For example, in *Plasmodium falciparum*, the aminopeptidases were suggested to be the enzymes to hydrolyze haemoglobin-derived peptides into amino acids (Gavigan et al., 2001). In *T. cruzi*, the major leucyl aminopeptidase (LAPTc) was suggested to function for

nutritional supply due to the lack of biosynthesis of some essential amino acid such as leucine (Cadavid-Restrepo et al., 2011). Based on the above results, proteolysis appears to play the critical role in order to initiate and establish the infection of *L. passim*. These genes include: Cytosol aminopeptidase family, catalytic domain containing protein, putative (Lp_000266800.1); glutamamyl carboxypeptidase, putative (Lp_000267300.1, Lp_000312800.1); GP63, leishmanolysin (Lp_000433700.1, Lp_100006400.1); proteasome alpha 2 subunit, putative (Lp_210013400.1); carboxypeptidase, putative (Lp_000297100.1); metallo-peptidase, Clan MF, Family M17 (Lp_270016200.1); dipeptidyl-peptidase 8-like serine peptidase (Lp_360020300.1) and hypothetical protein (Lp_000097400.1) with a Quick BLASTP hit of metallo-peptidase, Clan MA(E), Family M3 [*Leishmania mexicana* MHOM/GT/2001/U1103], E value=4e-61. Due to a wide range of substrate specificity (Bouvier et al., 1990) and the endopeptidase property of gp63 gene, *Lpgp63* might participate in the nutritional acquisition in the honey bee gut by cleaving long peptides into small peptides from the gut contents. Then, different exopeptidases including aminopeptidase and carboxypeptidase can cleavage these small peptide bond proximal to the amino (-NH2) or carboxy (-COOH) termini of the substrates. Thus, different endopeptidases and exopeptidases showed above may cooperate together in order to generate single amino acids for *L. passim* growth by protein hydrolysis. Moreover, many of the monoxenous trypanosomatids also release GP63-like enzymes into the extracellular medium. These include *Crithidia guilhermei* (de Melo et al., 2001;de Melo et al., 2002), *Crithidia deanei* (D'Avila-Levy et al., 2001;d'Avila-Levy et al., 2003), *Crithidia desouzai*, *Crithidia oncopelti*, *C. fasciculata* (D'Avila-Levy et al., 2001), *Crithidia luciliae*, *Leptomonas seymouri* (Jaffe and Dwyer, 2003), *Herpetomonas anglusteri, Herpetomonas roitmani* (Dos Santos et al., 1999), and *Blastocrithidia culicis* (Dos Santos et al., 2001;d'Avila-Levy et al., 2005). These observations imply the function of the GP63-like enzyme might involve in protein hydrolysis *in vitro*. Furthermore, these facts may also indicate a potential role of GP63 in the gut content degradation *in vivo*. This fact may also support the role of GP63 would be related to the nutrient acquisition. From the facts above, we may also support the hypothesis that the ancestral function of GP63 would be related to the survival of parasites in the dipteran hosts (Inverso et al., 1993).

3.3.5. *L. passim* gp63 may play critical roles for parasites to establish infection and colonize in the honey bee hindgut

My results on DEGs of *L. passim* during the infection also indicate that *Lpgp63* may play critical roles in honey bee guts. Firstly, when *L. passim* was ingested, honey bee expression of AMPs was triggered. The mechanism of AMPs is that these peptides disrupt the membrane potentials which lead to cell death by apoptosis (Haines et al., 2009). It has been reported that GP63 is also involved in the resistance to AMP killing. In *L. major*, the $GP63^{-/-}$ mutants (Joshi et al., 2002) were effectively killed by AMPs such as pexiganan. However, the *L. major* wild type cells were not. The AMPs were degraded by wild type *L. major* promastigotes metallopeptidase GP63 results in inactive fragments. This may lead to the inactivation of AMPs which further makes parasite resistance to AMPs killing (Kulkarni et al., 2006). Thus, *Lpgp63* may inactivate the honey bee AMPs and help them to escape from the honey bee innate immune surveillance. Secondly, after migrating to the honey bee hindgut, *Lpgp63* may function for nutrients acquisition by proteolysis as discussed above. In order to survive and proliferate in the honey bee hindgut, *Lpgp63* gene should always be up-regulated to fulfil the requirement of free amino acids from the guts. Meanwhile, *Lpgp63* could also function during the degradation of proteinaceous components within the honey bee hindgut. This process can provide room for *L. passim* to access to the cellular receptors on the hindgut cuticle cells. Thus, I could also assume *Lpgp63* promotes *L. passim* to migrate to the hindgut cells. Thirdly, after ingesting to the insect host, parasites need to bind to the specific gut cells in order to prevent being excreted from the faeces. This step may be regarded as an initial step to support the function of GP63 in the gut attachment. Also, this attachment step is critical for Leishmania to survive since it prevents excretion from the blood meals of sand flies (Kamhawi, 2006). In order to address the potential function of GP63 in the invertebrate guts, previous researchers investigated the role of GP63 in the relationship of mosquito *Aedes aegypti* and the monoxenic trypanosomatid *Herpetomonas samuelpessoai*. Other researchers demonstrated that *H. samuelpessoai* can successfully colonize in the *A. aegypti* gut and also *Herpetomonas* sp. is always found in this hematophagous mosquito (WEINMAN and Cheong, 1978;Corrêa-da-Silva et al., 2006;Pereira et al., 2010). Pre-treatment of *H. samuelpessoai* with metallopeptidases and anti-GP63 antibodies shows the binding ability of trypanosomatid was significantly affected *in vitro*.

Similarly, phospholipase C (PLC) treated parasites also showed a dramatical decrease of binding ability. This may also emphasize the involvement of these GPI-anchored molecules. Moreover, pretreated dissected insect guts with purified GP63 also induced a remarkable decrease in adhesion of parasites. This also indicates the saturation of gut cell receptors by GP63 metallopeptidases (de Melo et al., 2006). All these evidence indicate another role of GP63 in the invertebrate would be related to the attachment of parasites to the insect gut cells. Thus, *Lpgp63* may play a role in the hindgut attachment which prevents them excreted from honey bee faeces. Fourthly, more and more evidence is further reinforcing the critical role played by GP63 which significantly influence the vertebrate host cell signalling. Briefly, the function of GP63 on vertebrate cell signalling includes: 1) migration through the extracellular matrix, 2) evasion of complement-mediated lysis, 3) facilitation of promastigotes phagocytosis by macrophages, 4) resistance to antimicrobial peptide killing, 5) inhibition of natural killer cellular functions 6) degradation of macrophage and fibroblast cytosolic proteins with implications in cellular transduction signals and 7) promotion of survival of intracellular amastigotes (Yao et al., 2003;Yao, 2010;Isnard et al., 2012;Olivier et al., 2012). It can be seen from the GO terms that *Lpgp63* also alters the host's protein kinase-mediated signal transduction pathway and the NO-mediated signal transduction pathway. All these evidence indicate *Lpgp63* likely triggers the honey bee immune pathways. At last, by increasing the PTP activity in the host, *Lpgp63* negatively influences the JAK2/STAT immune pathways and also affect the production of NO. In fact, the GO terms of BP also revealed that *Lpgp63* is involved in the modulation of host signalling transduction pathway.

Figure 3.6 illustrates GO enrichment of up- and down-regulated genes at specific time points during the infection of honey bee hindgut. However, as for the down-regulated genes at day 12 PI and the up-regulated genes at day 27 PI, there is no GO enrichment with these genes. Basically, GO terms of the up-regulated genes uncovered the importance of carboxypeptidase and metallopeptidase activities as discussed above. For GO terms of the down-regulated genes, the respiratory chain and its related components dominate at the late stage of infection (20 PI and 27 PI). For example, they include mitochondrial electron transport, ubiquinol to cytochrome c (GO:0006122), mitochondrial respiratory chain complex III (GO:0005750), mitochondrial proton-transporting ATP synthase complex, catalytic core F(1) (GO:0000275), respiratory chain (GO:0070469) and mitochondrial ATP synthesis coupled electron transport (GO:0042775). From all the time points during infection,

it can be seen from the down-regulated GO terms that the protein synthesis/translation should be down-regulated in the infected *L. passim*. At day 7 PI, the top five GO enrichment of down-regulated genes are structural constituent of ribosome (GO:0003735)(*P-value*: 2.55E-32), cytosolic small ribosomal subunit (GO:0022627)(*P-value*: 3.23E-17), cytosolic large ribosomal subunit (GO:0022625)(*P-value*: 3.19E-12), cytoplasmic translation (GO:0002181)(*P-value*: 2.67E-07) and ribosomal small subunit assembly (GO:0000028)(*P-value*: 7.91E-07). Similarly, at day 20 PI, several top GO enrichment of down-regulated genes are structural constituent of ribosome (GO:0003735)(*P-value*: 3.05E-08), translation (GO:0006412)(*P-value*: 4.68E-07) and cytosolic small ribosomal subunit (GO:0022627)(*P-value*: 1.40E-05). There would be two potential reasons for this: 1) in the honey bee digestive tract, the amount of nutrients should be limited compared with the artificial medium, which could be the reason why the protein synthesis was decreased. 2) *L. passim* might decrease the expression of some surface proteins which help them to avoid the surveillance of the honey bee innate immune responses.

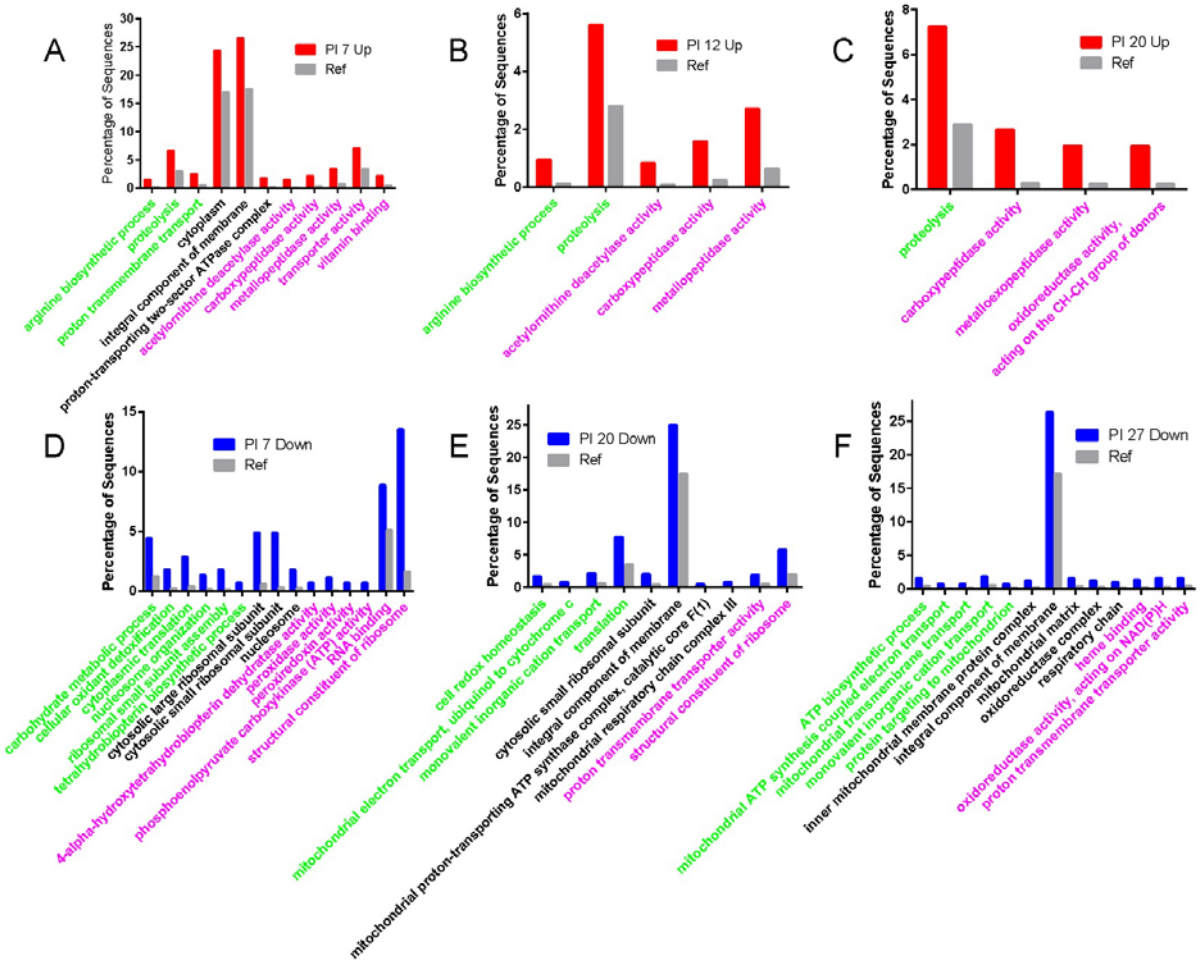


Figure 3.6 GO enrichment of DEGs of *L. passim* at four different time points during the infection. The DGEs were grouped

to up- and down-regulated ones for the GO enrichment test. GO enrichment terms were selected to the most specific one by FDR < 0.05. The up- and down-regulated genes were highlighted in red and blue, respectively. The reference test sets were coloured in grey. The GO terms are classified into three categories and highlighted in different colours on the X-axis which are green (Biological Process), black (Cellular Components) and magenta (Molecular Function). Y-axis represents the percentage of sequences. A, B and C: *L. passim* up-regulated genes at day 7, 12 and 20 PI. D, E and F: *L. passim* down-regulated genes at day 7, 20 and 27 PI. Up-regulated genes at day 27 and down-regulated genes at day 12 PI of *L. passim* have no GO enrichment terms.

3.3.6. Identification of DEGs of honey bee in response to *L. passim* infection

Figure 3.7 shows the number of up- and down-regulated honey bee genes are highest at 12 days after the infection. Specifically, 73 (40.1%) genes were up-regulated and 577 (87%) genes were down-regulated at 12 days PI among all honey bee DEGs in response to *L. passim* infection. These results indicate the honey bee's response to *L. passim* infection is most dramatic at 12 days PI. MDS plot and Pearson correlation heat maps were used to characterize the correlations between all RNA-seq samples, the reproducibility of duplicate, and the potential batch effects among all samples. Generally, the gene expression profiles of honey bee can be classified into 2 clusters except for the infected honey bee samples from day 12 PI. In addition, the infected samples from day 12 also distinct from each other which may indicate at this time points, the transcriptome profiles change quite fast. It is also possible that the differences between individuals might be the reason for this. All the honey bee samples from day 7 (infected and control) can be classified into one cluster. The rest of the honey bee samples can be classified into another cluster. The heat map from figure 3.7B also shows that the *L. passim* infected honey bee samples from day 12 were significantly different from all the other honey bee samples. The small amount of altered genes indicate that *L. passim* infection did not severely trigger the honey bee responses. Meanwhile, more RNA-seq can be performed near day 12 PI which may precisely record the gene expression profiles in detailed. At last, it should be pointed out some honey bee samples were mixed from different individuals in order to meet the quantity requirement. Individual honey bee under the natural condition may profile differently which may result in the distance between duplicate samples. The number of DEGs of honey bee in response to *L. passim* infection is not as many as *L. passim* adaption to the gut environment, i.e. only 835 genes were differentially expressed. Specifically, 182 genes were up-regulated whereas 653 genes were down-regulated at least once during *L. passim* infection. Gene expression profiles of honey bee are similar except the samples at 12 days after the infection. Also, I could conclude that

the both *L. passim* and honey bee transcriptomes changed dramatically at day 12 PI.

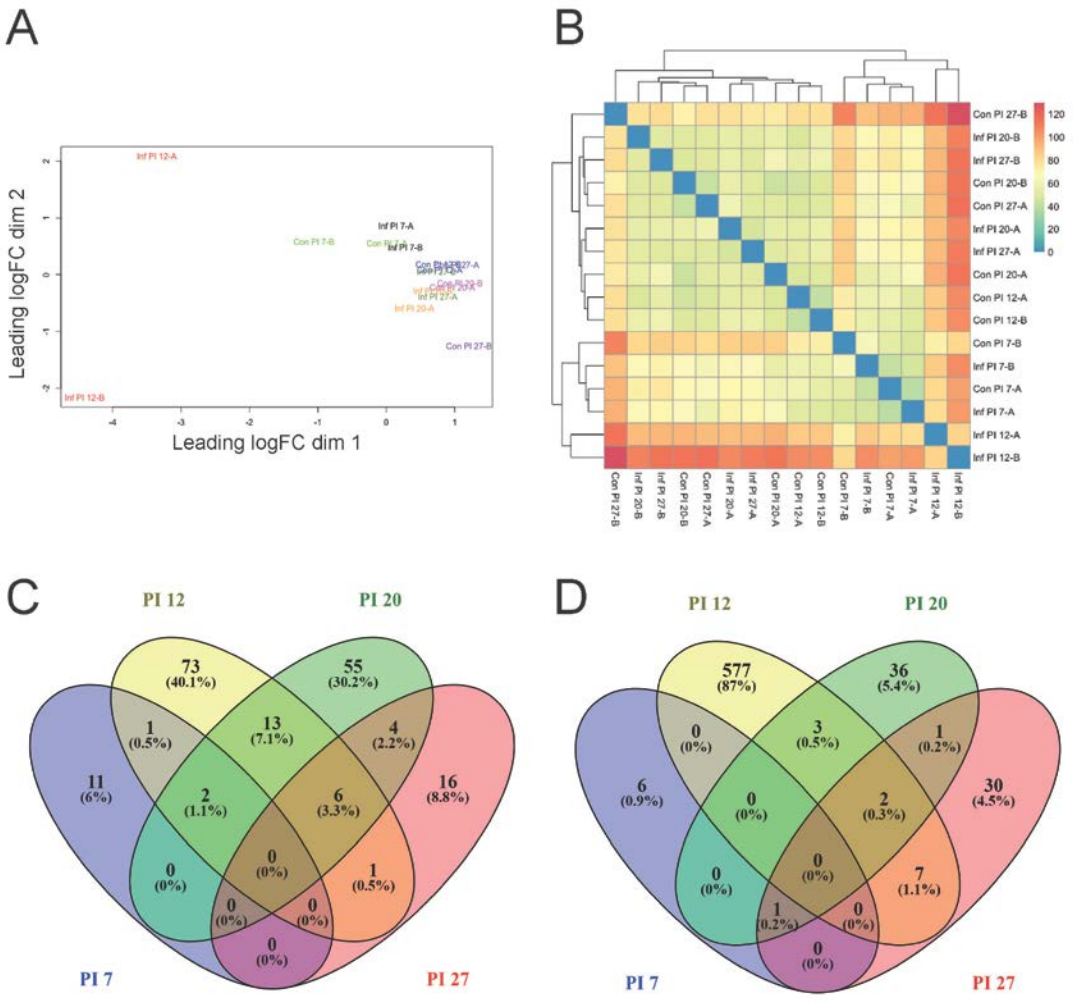


Figure 3.7 (A) MDS plot to illustrate a 2D projection of gene expression differences between samples. *L. passim*-infected (Inf PI 7-27) and the uninfected control (Con PI 7-27) samples at different time points during infection are indicated by different colors. Two replicates are marked as A and B. (B) Hierarchical analysis heat map using Pearson correlation. The levels of difference between samples are indicated by different colors. (C) Venn diagram to indicate honey bee genes up-regulated in response to *L. passim* infection at four different time points (PI 7-27) ($P < 0.05$). (D) Venn diagram to indicate honey bee genes down-regulated in response to *L. passim* infection at four different time points (PI 7-27) ($P < 0.05$).

3.3.7. GO enrichment of honey bee's DEGs in response to *L. passim* infection

Figure 3.7 shows that *L. passim* infection did not trigger extensive changes in honey bee gene expression profiles. Thus, only limited numbers of genes were used for the GO enrichment analysis. Specifically, Figure 3.8 shows GO enrichment terms can be retrieved with only four lists of the DEGs of honey bee in response to *L. passim* infection. On day 12 PI,

the down-regulated genes are enriched with GO-terms related to RNA processing and protein synthesis such as rRNA processing (GO:0006364) (*P-value*: 5.81E-07), mRNA processing (GO:0006397) (*P-value*: 3.87E-05), RNA polymerase complex (GO:0030880) (*P-value*: 3.10E-04), preribosome (GO:0030684) (*P-value*: 5.08E-04), and RNA splicing (GO:0008380) (*P-value*: 6.43E-04). During bacterial infection, it has been shown the protein translation was inhibited in fruit fly (Chakrabarti et al., 2012). Figure 3.8 shows that the majority of GO terms enriched with the up-regulated genes on day 12 PI are related to mitochondrial activity and ETC. These include: proton transmembrane transport (GO:1902600) (*P-value*: 1.47E-15), NADH dehydrogenase (ubiquinone) activity (GO:0008137) (*P-value*: 1.55E-12), mitochondrial electron transport, NADH to ubiquinone (GO:0006120) (*P-value*: 1.73E-11) and respiratory chain complex IV (GO:0045277) (*P-value*: 9.53E-05). These results indicate that the infected honey bee stimulates mitochondrial activity compared with the uninfected control bee. This may indicate that *L. passim* infection could alter the honey bee energy homeostasis since the ultimate product of ETC is ATP. On day 20 PI, the enriched GO terms are all related to phagocytosis such as phagocytosis, recognition (GO:0006910) (*P-value*: 1.25E-05) and positive regulation of phagocytosis (GO:0050766) (*P-value*: 1.25E-05). These GO terms may indicate that phagocytosis play role in infected honey bee in response to *L. passim* infection.

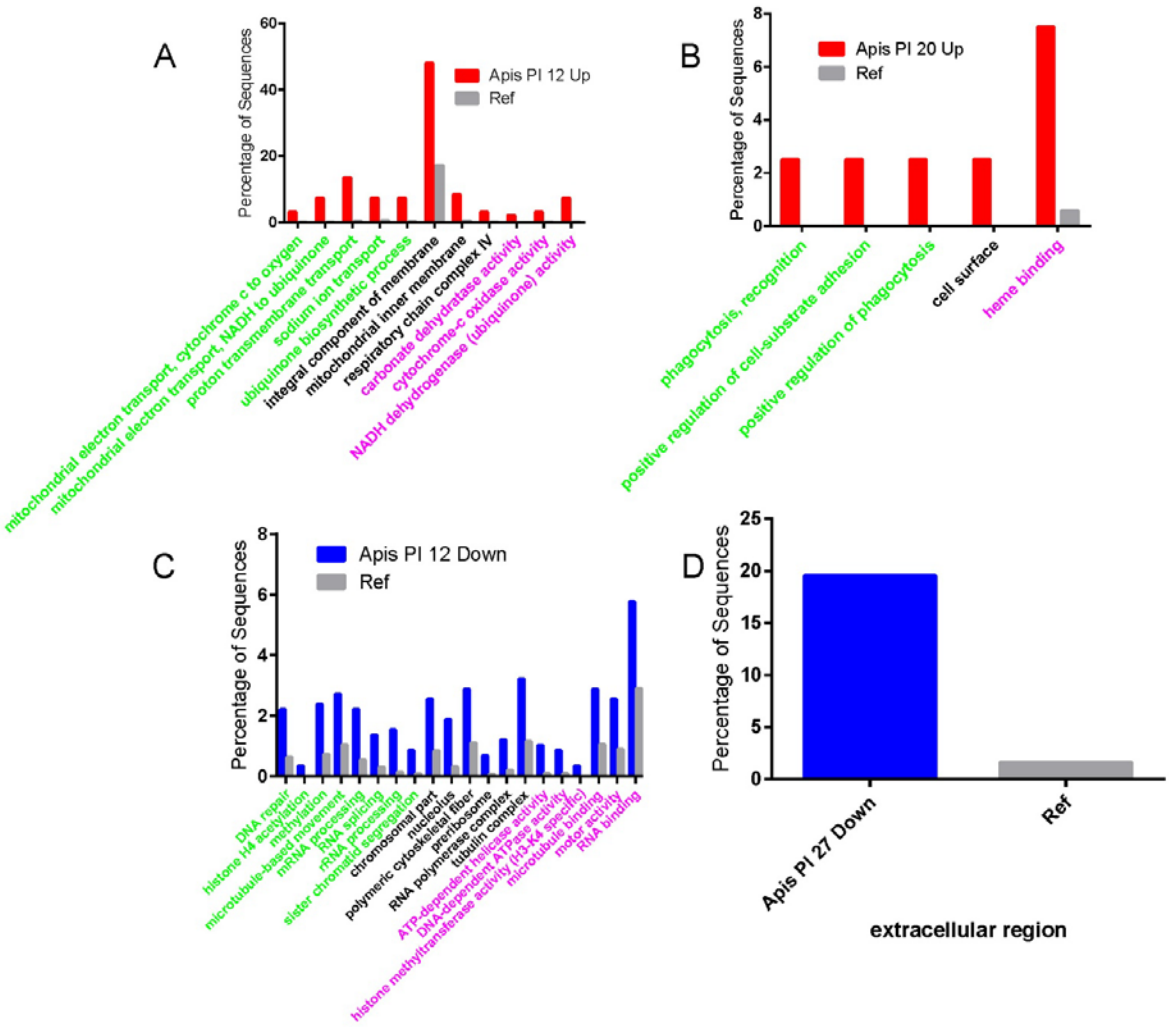


Figure 3.8 GO terms enriched with honey bee genes in response to *L. passim* infection. The DEGs of honey bee were divided to up- and down-regulated genes for the GO enrichment tests. The enriched GO terms were selected to the most specific ones by FDR < 0.05. The up- and down-regulated genes are highlighted in red and blue, respectively. The reference test sets are coloured in grey. The GO terms are classified into 3 categories and highlighted in different colours on the X-axis; green (Biological Process), black (Cellular Components) and magenta (Molecular Function). Y-axis represents the percentage of sequences.

3.3.8. *L. passim* infection triggers the immune response of honey bee

The expression of honey bee AMPs were regulated by Toll and Imd pathways. Specifically, a family of pathogen recognition receptors (PRRs) bind to the pathogen associated molecular patterns (PAMPs) of different pathogens such as fungi and bacteria which initiate the Toll pathway. After PAMP binding, the serial of proteolytic cascade cleavage the pro-Spaetzle into mature Spaetzle which can bind to the Toll receptors that anchored on the cell membrane (Brutscher et al., 2015). There are two putative Spaetzle orthologues which can

be found in the honey bee genome (Evans et al., 2006). Then, the dimerization of Toll on the membrane results in the assembly of dMyD88, Tube and Pelle. Pelle is probably participating in the degradation of NF-kappa B inhibitors such as Cactus-1, -2 and -3. Later, this resulting in the release of several transcription factors such as Dorsal-1A and -1B. Ultimately, the nuclear translocation Dorsal increases the expression of AMPs. Another important pathway in the honey bee immune system is the Imd pathway which also regulates the expression of AMPs. The activation of Imd pathway is initiated by the binding of peptidoglycan (PGN) to the peptidoglycan recognition protein LC (PGRP-LC) on the cell membrane (Brutscher et al., 2015). In fruit fly, PGN especially from Gram-negative bacteria were often observed to trigger the Imd pathway with the function of recognition (Evans et al., 2006). Also, the PGRP-LC in *Drosophila* has three alternative splice forms which carry different PGN recognition domains. This phenomenon likely indicates the wide specificity of fruit fly recognition in Imd pathway (Werner et al., 2003;Mellroth et al., 2005;Chang et al., 2006). Then, the binding of PGN and PGRP-LC activate the Immune deficiency (IMD) which is an adaptor protein. Later, Relish was phosphorylated by IKK protein complex. In the final step, the Relish was cleavage by caspase Dredd (Death-related ced-3/Nedd2-like) and the Relish regulates the expression of AMPs (Brutscher et al., 2015). Interestingly, the honey bee genome also codes all orthologues components of Imd pathway which indicates the Imd pathway is highly conserved in both honey bee and fruit fly (Evans et al., 2006). The proposed honey bee Toll and Imd immune pathways in response to *L. passim* infection were shown in Figure 3.9. Not surprisingly, *L. passim* infection triggers the immune responses of the honey bee at the early stage of infection. Specifically, there are two AMPs, *hymenoptaecin* and *defensin-1*. These two genes were found to be up-regulated on day 7 PI. Moreover, honey bee *defensin-1* was up-regulated across the entire infection period except for day 27 PI. The up-regulation of AMPs related genes is widely reported when the pathogens invaded to the invertebrate host (Boulanger et al., 2001;Hao et al., 2003;Boulanger et al., 2004;Boulanger et al., 2006;Antúnez et al., 2009;Riddell et al., 2009;Aufauvre et al., 2014;Brutscher et al., 2015;Brutscher et al., 2017). However, the expression of honey bee *abaecin* was suppressed by *L. passim* infection in the early stage of infection (7 days PI). Similarly, bumble bee immune genes *MyD88* and *Relish* were also differentially expressed when infected by *C. bombi* by RT-PCR (Schlüns et al., 2010). In addition, by using RNA-seq, it has been shown that several bumble bee genes related to signalling pathways were differentially expressed in

response to *C. bombi* infection such as *Spatzle*, *MyD88* and *Dorsal*. Unsurprisingly, *C. bombi* infection also triggers the expression of several AMPs. These genes include *defensin*, *hymenoptaecin* and *apidaecin* (Riddell et al., 2014). These facts suggest that *Crithidia* also involved in the immune pathways in response to infection in bumble bee. Also, bumble bee shares high similarity in terms of Toll and Imd pathways. Consistent with the previous research, my RNA-seq did not show any change on the expression of *apidaecin* at four time points (Schwarz and Evans, 2013). From my RNA-seq result, honey bee *NF-kappa-B inhibitor cactus 1* was up-regulated at day 7 PI. The Cactus-1 functions as a negative factor in both Toll and Imd pathways. The degradation of these inhibitors results in the release of transcription factors such as Dorsal (Toll pathway) and Relish (Imd pathway). Previous research demonstrated that after silencing *Relish*, the *abaecin* and *hymenoptaecin* were affected which indicates that the expression of these AMPs is controlled by Imd pathway. However, it also shows that *defensin-1* was not affected by *Relish* silencing which indicates the expression of *defensin-1* might be controlled by Toll pathway (Schlüns and Crozier, 2007). More recently, it has been shown that both *Dorsal* and *Relish* are involved in the expression of *abaecin*. In addition, the transcription factor *Dorsal* regulates the expression of *defensin-1* and *Relish* is involved in the expression of *hymenoptaecin* (Lourenço et al., 2013). Moreover, it should be pointed out that the expression of AMPs is not only controlled by infectious stresses but also can be induced by other non-infectious stresses. Such as aseptic wounding (Erler et al., 2011), temperature and starvation (Tsuzuki et al., 2012). Moreover, it also shows that over-expressing growth-blocking peptide increases the AMP expression in the fruit fly. This phenomenon also implies that the expression of AMPs does not require the participation of immune pathways (Tsuzuki et al., 2012). Thus, the expression of AMPs can be induced by different factors. Herein, my RNA-seq result indicates the Toll and Imd pathways were likely affected by *L. passim* infection. However, how different AMP families were regulated upon *L. passim* and the specific mechanisms in the signalling pathways need to be elucidated in the future. Immune responses cost a lot to the host, and thus, the activation of immune stress by *L. passim* may also alter the behaviours of infected bees which may ultimately affect the fitness of the hive.

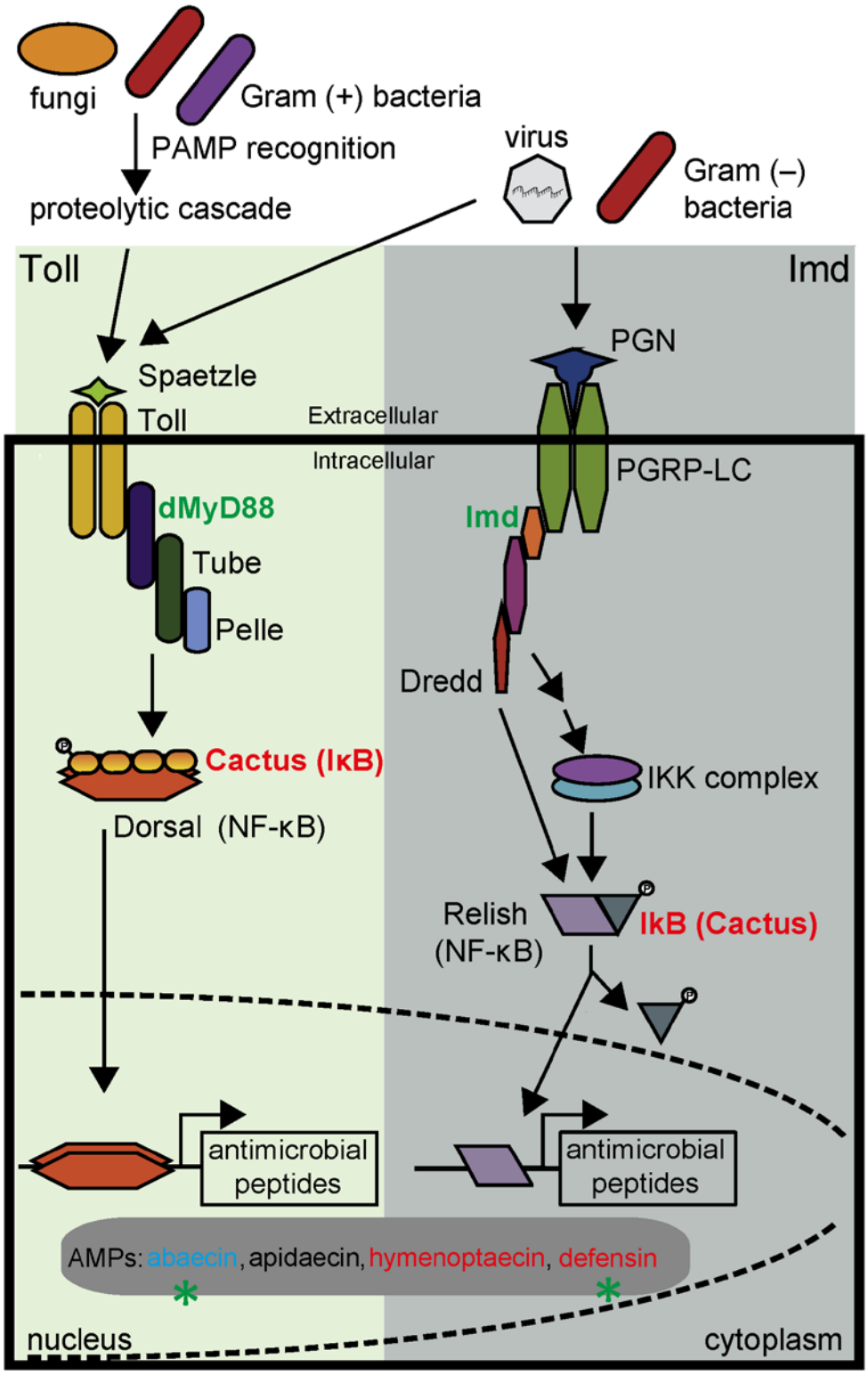


Figure 3.9 Honey bee Toll and Imd immune pathways in response to *L. passim* infection. Several key components were shown and genes participate in the responses to *L. passim* were coloured in different font. Green font and asterisk indicate the genes altered by *L. passim* challenge from the previous research (Schwarz and Evans, 2013). Red font and blue font indicate the up- and down-regulation genes by *L. passim* infection according to the RNA-seq result. Adapted and modified from (Brutscher et al., 2015)

3.3.9. *L. passim* infection significantly alters the energy metabolism of honey bee

Interestingly, on day 12 PI, *L. passim* triggers a dramatic change in the mitochondria activity of honey bee gut cells since basically all of the GO terms are related to the mitochondrial functions. At this time point, a significantly up-regulated of ETC indicate a strong demand for energy. *L. passim* may consume short-chain fatty acids (SCFAs) produced by hindgut microbiome. Uptake of trehalose appears to be stimulated in the *L. passim* infected honey bee gut epithelial cells by increasing several trehalose transporters (*PREDICTED: facilitated trehalose transporter Tret1-like*) gene expression. This can be seen at 7, 12, and 20 but not 27 days PI. Lack of SCFA supply from the gut lumen may stimulate the uptake of trehalose from the hemolymph. Some of the up-regulated genes by *L. passim* infection are related to the lipid and fatty acid metabolism. This change can be found on day 12 PI and the later time points. These genes include: apolipoprotein D-like, elongation of very long chain fatty acids protein AAEL008004-like, phospholipase A1 and fatty acyl-CoA reductase 1-like. Thus, *L. passim* infection appears to modify the host lipid and fatty acids metabolism.

3.3.10. *L. passim* infection triggers the peritrophic membrane (PM) re-organization of infected honey bee

The PM is also known as the peritrophic matrix which separates the ingested food from the intestinal epithelium. PM can be divided into two different types according to the site of its synthesis (Peters, 2012). Take the hematophagous insect as an example, type 1 PM is produced from midgut epithelial cells in the response to direct blood feeding. This response results in the formation of a bag-like structure that completely surrounds the ingested meal. Type 2 PM is produced by a specialized organ which is near the junction between foregut and midgut called cardia (Jacobs-Lorena and Oo, 1996;Tellam et al., 1999). The type 2 PM is a thin open-ended tube-like structure lining not only the midgut but also the hindgut. Both PM 1 and PM 2 contain chitin and glycoproteins (Shao et al., 2001). The PM facilities digestion and also forms a physical protective barrier to prevent the invasion of ingested pathogens (Lehane, 1997;Kuraishi et al., 2011). On day 27 PI, chitinase-3-like protein 1 gene and peritrophin-1-like gene were up-regulated. The chitin fibrils of the PM are assembled into a wide cross-hatched pattern connected by peritrophin (Dinglasan et al., 2009). Similarly,

bumble bee also shows the PM response responds to *C. bombi* infection (Riddell et al., 2014). I could assume that *L. passim* infection triggers the repair or re-organization of PM in infected bee guts.

3.3.11. *L. passim* infection induces the down-regulation of honey bee odorant binding proteins

On day 20 PI, four genes coding odorant binding proteins (*Obp21*, *Obp16*, *Obp4*, *Obp17*), were down-regulated compared with the control. In honey bee, 21 OBP genes were identified. Some OBP proteins are expressed exclusively in the antennae of adult worker whereas the others are expressed ubiquitously in the body. They are also highly expressed in the cuticle (Forêt and Maleszka, 2006). These OBP proteins recognize and bind a variety of airborne molecules but also participate in the transportation of some hydrophobic cuticular compounds (Shanbhag et al., 2001). More recently, the expression of odorant binding proteins is related to the hematopoiesis which induced by the gut symbiotic bacteria, *Wigglesworthia* (Benoit et al., 2017). It has been shown that the expression of Obp6 in tsetse flies and its orthologue Obp28a in fruit flies have the same function. Moreover, the Obp6 and Obp28a were also controlled by the symbiont. These odorant binding proteins control the generation of a specific type of hemocyte called crystal cells in larvae during the development stage. The crystal cells are responsible for the production of melanin. Melanin involved in the killing of pathogenic microbes and preventing the loss of hemolymph from the wounds in the insect exoskeleton (Benoit et al., 2017). In both tsetse and fruit fly larvae, they express a reduced level of odorant binding protein which inhibits the melanization pathway which increases the mortality after wounding (Benoit et al., 2017). It can be seen that odorant binding proteins participate not only in recognition of airborne molecules, but also play a critical role in the immune response pathway in different invertebrates. Thus, in the infected honey bees, the immune system is likely affected by *L. passim* infection. However, more experiments related to honey bee odorant binding proteins should be done in the future to elucidating the mechanisms.

3.3.12. *L. passim* infection induces the expression of honey bee *takeout* gene in the hindgut

Most interestingly, in *L. passim* infected honey bee, *takeout* or *takeout-like* genes were significantly up-regulated compared to the uninfected control bees at the later stage of infection. These *takeout* genes were up-regulated in the *L. passim* infected honey bee on day 20 and 27 PI. In fruit fly, it is mainly expressed in the head whereas it is undetectable under normal condition in the adjacent gastrointestinal tract regions. In fact, *takeout* mRNA increased by starvation in both brain and body tissues (Sarov-Blat et al., 2000). The expression of *takeout* gene in gastrointestinal tract regions was specifically induced by starvation (Sarov-Blat et al., 2000). Later on, the *Drosophila takeout* gene has been reported to be an essential component to maintain the energy homeostasis. In the *takeout* mutant fruit flies, they consume more food compared than WT files. Furthermore, the increase in food intake is dramatically lower in *takeout* mutant files compared to WT files. *Takeout* gene also involved in the foraging activity. Specifically, in the WT flies, the locomotor activity increase when the starvation gradually happens (Martin, 2003;2004). However, under such circumstance, *takeout* mutant does not show a dramatic increase in locomotor activity (Meunier et al., 2007). In the sand fly, silencing of *takeout* gene promotes an increased propensity on blood-feeding which may also indicate the *takeout* gene is an important gene controlling blood-feeding activity (Das and Dimopoulos, 2008). All these facts indicate that the *takeout* gene is related to starvation and may consequently affect other behaviours in different organisms. Since the samples for RNA-seq analysis were collected from honey bee rectum, the up-regulation of honey bee *takeout* gene also indicates that the infected bees are likely to be under starvation. As observed in the fruit fly, expression of honey bee *takeout* would also increase in the brain by *L. passim* infection.

3.3.13. *L. passim* infection alters the exon usage of infected honey bee

Using DEXSeq package, I performed an alternative splicing analysis using “condition” as a model (FDR < 0.05) at four different time points during the infection. On day 7, 12, 20 and 27 PI, there are 10, 1335, 538 and 332 loci displayed alternative splicing depending on “infected” and “uninfected” states. Many genes show differential exon usage due to the interaction of *L.*

passim infection at different time points. The complete output of DEXSeq analysis can be found in the supplementary document. Figure 3.10 shows the different exon usage between *L. passim*-infected honey bee and control honey bee. Only *titin-like* (LOC554319, GB40010) gene was alternatively spliced across the entire infection period. *Twitchin* is a member of the *titin* gene family which produces filamentous proteins that participate in the mechanical transduction signal in muscle (Mayans et al., 2013). Previous experiments showed the differential exon usage of *Titin* when Plasmodium or bacteria infect mosquito (Garver et al., 2008). Bumble bee *twitchin* gene was also shown a dramatical alternative splicing when infected by *C. bombi* (Riddell et al., 2014). Although *L. passim* infection results in changing alternative splicing patterns in many genes, the functions of these isoforms need to be further investigated. *Titin* and *Twitchin* are good candidates to be explored in future.

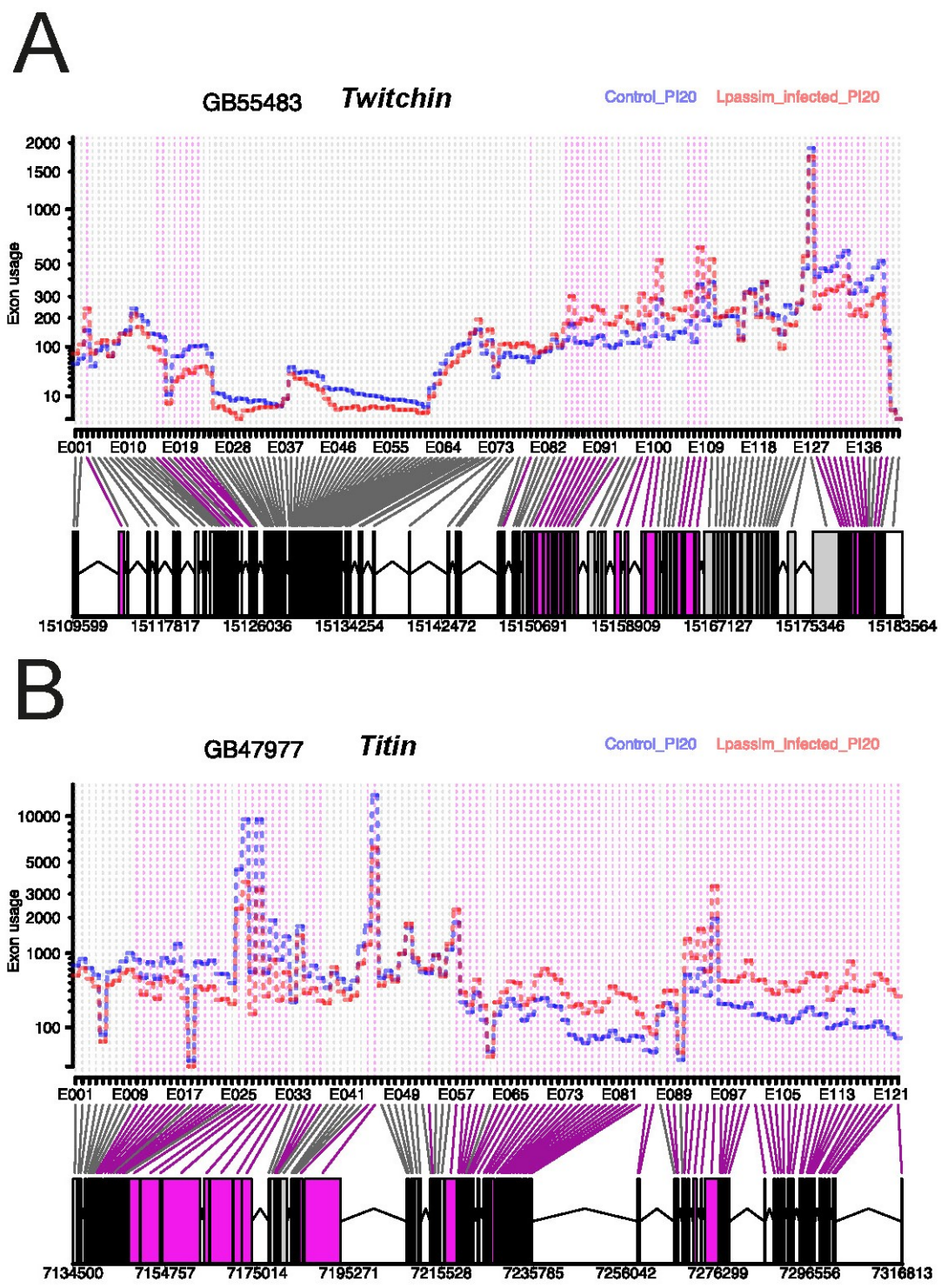


Figure 3.10 Differential exon usage of *Twitchin* (A) and *Titin* (B) on day 20 PI. DEXSeq removes the gene-level changes in expression so the differential exon usage is clear. The gene models are shown below. The purple blocks represent the exons differentially spliced between *L. passim* infected and control honey bees. The blue line indicates the control honey bee sample whereas the red line indicates the infected honey bee sample.

3.4. Conclusions

In this part of the project, I present a comprehensive transcriptomic analysis with this trypanosome, *L. passim*. Concurrently, how *L. passim* adapts to honey bee hindgut microenvironment, and more importantly, how *L. passim* potentially affects honey bee host was proposed here. By using RNA-seq, it becomes possible to understand the effects of *L. passim* infection on honey bee and then led me to test the changes of Vg in the fat body of honey bee by *L. passim* infection. Thus, there is no doubt that RNA-seq is truly a powerful tool in terms of understanding the host-parasite interaction, especially when the phenotypes of infection were not completely understood.

Like *Leishmania* and *Trypanosoma* spp., *L. passim* also needs to develop unique adaptive mechanisms in order to survive in honey bee hindgut. After escaping from the immune surveillance of honey bee, *L. passim* reaches to the hindgut of honey bee and starts to colonize there. A large number of genes are differentially expressed when *L. passim* establishes the infection and colonizes in the honey bee hindgut. *L. passim* proteins with carboxypeptidase activity, metallopeptidase activity and proteolysis are continuously up-regulated during the infection. As a member of trypanosomatid, my results also emphasize that gp63 may play essential roles for *L. passim* infection. Also, the down-regulation of *LpTXNPx* across the entire infection period may help to directly enhance the binding of UMSBP to UMS in order to stimulate kinetoplast DNA replication at the early phase of the infection and adapt to the anaerobic condition of honey bee hindgut.

For the response to *L. passim* infection, honey bee immune responses such as the activation of AMPs are induced. This phenomenon is consistent with the response of bumble bee against *Crithidia* (Riddell et al., 2009;Riddell et al., 2011;Deshwal and Mallon, 2014;Riddell et al., 2014). Also, the expressions of several odorant binding proteins were reduced in the *L. passim* infected honey bees which may indicate the bee immune response to *L. passim* infection. Moreover, since the expression of honey bee AMPs was controlled by Imd and Toll pathways (Evans et al., 2006;Brutscher et al., 2015), the alteration of different AMPs and *NF-kappa-B inhibitor cactus-1* in the *L. passim* infected honey bee highly implies that several immune pathways were also involved in response to *L. passim* infection.

Most interestingly, I found a significant change in the mitochondrial activity in the infected honey bee. Moreover, in the late stage of infection, several fatty-acid related genes were

also differentially expressed. This may indicate the fatty acid metabolic pathway was affected by *L. passim* infection.

Surprisingly, only a small number of genes were up-regulated during *L. passim* infection. Also, there is no gene up- or down-regulated across the entire infection period. This result shows that the honey bee responses to *L. passim* infection periodically changes. My data show *L. passim* infection slightly increase the mortality under laboratory condition since no pollen was provided. The previous report in Belgium indicates a correlation of *L. passim* infection with winter mortality of honey bee colony (Ravoet et al., 2013). My data here may support that *L. passim* would be a new threat to honey bees when the food is limited (e.g. during the over-wintering period). In the late stage of *L. passim* infection, *takeout* gene was highly expressed, suggesting that honey bee is under starvation.

There is no report indicating that *C. bombi* infection induces starvation of bumble bee. I could assume during the long-term infection, *C. bombi* may also alter the nutritional state of bumble bees since they are close species. With the help of RNA-seq, my results have given insights into how this trypanosomatid interacts with its natural host. However, there are more works to do in order to understand the specific function of genes during infection as well as to verify the hypothesis I proposed here.

4. Chapter IV

Establishment of CRISPR/Cas9 system in *L. passim*

4.1. Introduction

4.1.1. Recent CRISPR/Cas9 system applied to trypanosomatid

RNA inference (RNAi) was discovered two decades ago which is a powerful tool to study gene function (Fire et al., 1998). However, in trypanosomatids, RNAi machinery from the majority of them are missing and knocking down gene functions have been successful with only a few species (Docampo, 2011;Kolev et al., 2011). Moreover, using homologous recombination (HR) to KO endogenous genes usually is time-consuming and the second round of transfection is needed (Santrich et al., 1997). These facts led to the urgent demand for an effective molecular tool to manipulate the genome in order to study the gene functions in these medically important species. The first CRISPR/Cas9 system in Leishmania was established in *L. major* in 2015 (Sollelis et al., 2015b). In this paper, a Cas9 expression plasmid was firstly transfected to obtain the promastigotes which constitutively express Cas9. Then, a 6kb circular plasmid which containing sgRNA (driven by U6 promoter) and donor DNA containing puromycin resistance gene and long homologous arms were transfected to the Cas9-expressing *L. major*. More importantly, this paper also confirmed the off-target effect is not present by using RNA-seq. Later, the CRISPR/Cas9 system was established in *L. donovani* by using miltefosine transporter (LdMT) as a target under the negative selection of miltefosine (Zhang and Matlashewski, 2015). In this paper, for the sgRNA expression vector, hepatitis delta virus (HDV) ribozyme immediately downstream of sgRNA sequence can increase the KO rate. The function of HDV ribozyme is to precisely terminate the sgRNA transcription. Furthermore, a vector which could express two sgRNAs simultaneously was used to induce large deletion in the endogenous gene as expected. Using this two-sgRNAs expression vector, the efficiency to knock-out gene increased four-fold compared to the single sgRNA expression vector. At last, it has been shown that *L. donovani* uses MMEJ pathway to repair the DSB in the absence of homologou s donors. If the homologous donor is present, HDR is preferred compared with MMEJ. Furthermore, the same group also used the CRISPR/Cas9 system they established in *L. donovani* to demonstrate the patterns of chromosomal translocations by introducing homologous flanking oligonucleotide (Zhang et al., 2017).

In 2015, the CRISPR/Cas9 system was first applied to KO several exogenous and endogenous

genes in *T. cruzi* (Peng et al., 2015). In this paper, the author transfected the Cas9 expressing cell line with sgRNA synthesized *in vitro*. This method results in a low gene KO rate when using tdTomato homologous flanking arms to replace the eGFP gene. This also indicates the *T. cruzi* can use HDR to repair DSB in the presence of homologous donors. As for endogenous genes, without homologous donor, the author shows that *T. cruzi* could use MMEJ to repair the DSB. Also, the cytotoxicity of Cas9 was observed. Later, Docampo's group used a different approach to improve the CRISPR/Cas9 systems in *T. cruzi* (Lander et al., 2015a). They used one-vector or two-vector strategy to express Cas9 and sgRNA. More importantly, DNA donor containing drug resistance marker was provided in order to facilitate the homologous recombination. This approach reported here increased genome editing efficiency to 100% with a clean gene KO in only 5 weeks after co-transfection. Furthermore, the same group used this system to study several endogenous gene functions in this species (Lander et al., 2016b;Chiurillo et al., 2017). In 2017, a detailed protocol to tag the C-terminus of endogenous protein in *T. cruzi* was well-established (Lander et al., 2017).

In *T. brucei*, the RNAi machinery is present and RNAi can be used to study the gene functions. Also, it has been shown that the classical HR method could effectively KO the endogenous genes. However, not surprisingly, this procedure takes two rounds of transfection which is time-consuming (Santrich et al., 1997;Docampo, 2011). More recently, in *T. brucei* bloodstream form, CRISPR/Cas9 has also been applied to rapidly modify or delete two alleles of a target gene in just one step (Beneke et al., 2017). Also, in this paper, *L. major* and *L. mexicana* expressing Cas9 and T7 RNA polymerase were generated first. Then, cassettes containing homologous arms and sgRNAs under T7 promoter were transfected to these cell lines. More importantly, this paper also shows that CRISPR/Cas9 system can significantly reduce the length of homologous flanking arms compared with the classical HR method in *L. mexicana*. Moreover, by using two specific plasmids for endogenous tagging and KO, this toolkit together with the online primer design website dramatically decrease the time needed to KO endogenous genes. An improved CRISPR/Cas9 system was reported to express tetracycline inducible Cas9 and sgRNA into *T. brucei*. The expression of Cas9 protein and sgRNA was under the control of tetracycline which can control the time of genome editing. Also, this inducible CRISPR/Cas9 system make the single base editing possible without the detectable toxicity of Cas9. By applying this new CRISPR/Cas9 system in *T. brucei*, the successful genome editing rate almost reached 100%. Moreover, this inducible system can

modify both alleles of the target gene without using drug resistance markers. Furthermore, the KO cells can be obtained in a week which significantly decreases the time used for KO endogenous genes compared with the previous methods as shown above (Rico et al., 2018).

4.1.2. Repair DSB pathways induced by CRISPR/Cas9 in trypanosomatid

4.1.2.1. Homology-Directed Repair (HDR)

In kinetoplastida, they have to repair the DSB induced by CRISPR/Cas9 system by two repair mechanisms. Namely, they are homology-directed repair (HDR) and microhomology-mediated end joining (MMEJ). HDR uses homologous recombination (HR) to repair the DBS on the genome which is critical for any organisms to maintain the genome stability. Most of the HR machinery is highly conserved in eukaryotic cells. In trypanosomatid parasites, the majority of the enzymes involved in HR is conserved too (Kelso et al., 2017b). Recently, it has been widely reported that *T. cruzi*, *T. brucei*, *L. donovani*, *L. major* and *L. mexicana* prefer to use HDR to repair the DSB generated by CRISPR/Cas9 when the homologous donor is present (Lander et al., 2015a;Peng et al., 2015;Sollelis et al., 2015a;Zhang and Matlashewski, 2015;Lander et al., 2016b;Beneke et al., 2017;Chiurillo et al., 2017;Lander et al., 2017;Zhang et al., 2017;Rico et al., 2018).

There are several key factors which have been shown to participate in the HR process. For example, in *T. brucei*, one of the major factors is RAD51 (radiation sensitive 51), an enzyme involved in DNA break repair. Moreover, *TbRAD51* was shown to be essential when the parasite responses to DNA damage and the DSB repair pathway by HR also shown to be the predominant DSB repair pathway (Barnes and McCulloch, 2007;Glover et al., 2008). As for the efficiency, when the homologous DNA donor is present, *T. cruzi* could use the HR to repair the DSB with high efficiency (Lander et al., 2015a). All these evidence indicate that HDR acts an essential pathway in the responses to DNA damage as well as the adaption to the environment.

4.1.2.2. Microhomology-Mediated End Joining (MMEJ)

Another important DSB repair pathway in trypanosomatid is MMEJ. *T. cruzi* and *L. donovani* use the MMEJ pathway to repair the DSB induced by CRISPR/Cas9 in the absence of homologous DNA donors (Peng et al., 2015;Zhang and Matlashewski, 2015). However, HR

inhibitors were applied to inhibit the activity of RAD51 in *L. donovani* to suppress the HR pathway and enhance the MMEJ pathway. They proposed this inhibition would increase MMEJ pathway which might be more efficient to study the gene function since no drug selection is needed. However, the results show no increase of MMEJ rate, indicating HDR did not play a dominant role as expected. It has been reported that in *T. brucei*, GC-content near the DSB together with the proportion of matched base pairs promote the MMEJ pathway (Glover et al., 2010). In yeast, after elimination of RAD51, it showed an increased MMEJ frequency, suggesting yeast could switch to MMEJ in the absent of HR (Villarreal et al., 2012;Deng et al., 2014). The MMEJ pathway contains several key steps which are: 1) End resection 2) Micro-homology annealing 3) Fill-in synthesis 4) Heterologous flaps removal and 5) Ends sealing. Also, there are several key factors which play critical roles in the MMEJ pathway such as cyclin-dependent kinase (CDK), DNA polymerase θ and Mre11 (Sfeir and Symington, 2015). Both HR and MMEJ were initialized by the end resection. Afterwards, how the cells select the DSB repair pathway has been investigated. The DSB repair pathway selection was related to cell cycle and end resection. Cells prefer to use HR to repair DSB during S-G2 phase since during this cell phase, the sister chromatin is present which can be used as a template for HR. During G1 phase of the cell cycle, MMEJ pathway is preferred to be selected by cells since the resection activity was low (Aylon et al., 2004;Ira et al., 2004). Since MMEJ pathway uses the micro-homologies to anneal resection end on the genome, MMEJ always introduces indels near the cleavage site which is always error-prone. Compared with HR, it uses the sister chromatin to repair the DSB which make this repair method is error-free. Originally, MMEJ pathway was regarded as a backup pathway to repair the DSB. However, the recent studies indicate that MMEJ as a major DSB pathway for *L. donovani* to repair the DNA damage (Zhang et al., 2017). To date, the physiological function of MMEJ is still unclear. Also, there is no method available to control the micro-homologies usage by organisms in order to control the indels patterns. We could assume that MMEJ increases the genome variability.

Due to the successful establishment of CRISPR/Cas9 in various trypanosomatid species recently, I intended to establish CRISPR/Cas9 system to understand the gene functions in *L. passim*. After establishing the CRISPR/Cas9 genome engineering system in *L. passim*, I could identify and characterize the genes necessary to establish the infection in honey bee gut.

4.2. Materials and Methods

4.2.1. Plasmids constructs

The following plasmids were bought from addgene website (http://www.addgene.org/) in order to obtain different *L. passim* cell lines and establish CRISPR/Cas9 system by 1V (one vector) and 2V (two vectors) in *L. passim*: pTrex-Neo-tdTomato (Canavaci et al., 2010) (Plasmid #47975), pUC57-sgRNA expression vector (Shen et al., 2014) (Plasmid #51132), pTrex-b-NLS-hSpCas9 (Peng et al., 2015) (Plasmid #62543), pTrex-n-eGFP (Peng et al., 2015) (Plasmid #62544), pUC_sgRNA (plasmid #68710), Cas9/pTREX-n (Lander et al., 2015b) (Plasmid #68708), pSPneogRNAH (plasmid #63556), pLPhygCAS9 (Zhang and Matlashewski, 2015) (plasmid #63555). pSpCas9n(BB)-2A-Puro (PX462) V2.0 (Ran et al., 2013)(Plasmid #62987). Plasmid pCsV1300 was a gift from Yutong LIU from Northeast Normal University. Specifically, pTrex-Neo-tdTomato and pTrex-n-eGFP were used as fluorescent markers to increase and monitor the electroporation efficiency of *L. passim*. The plasmid constructs are shown in Figure 4.1.

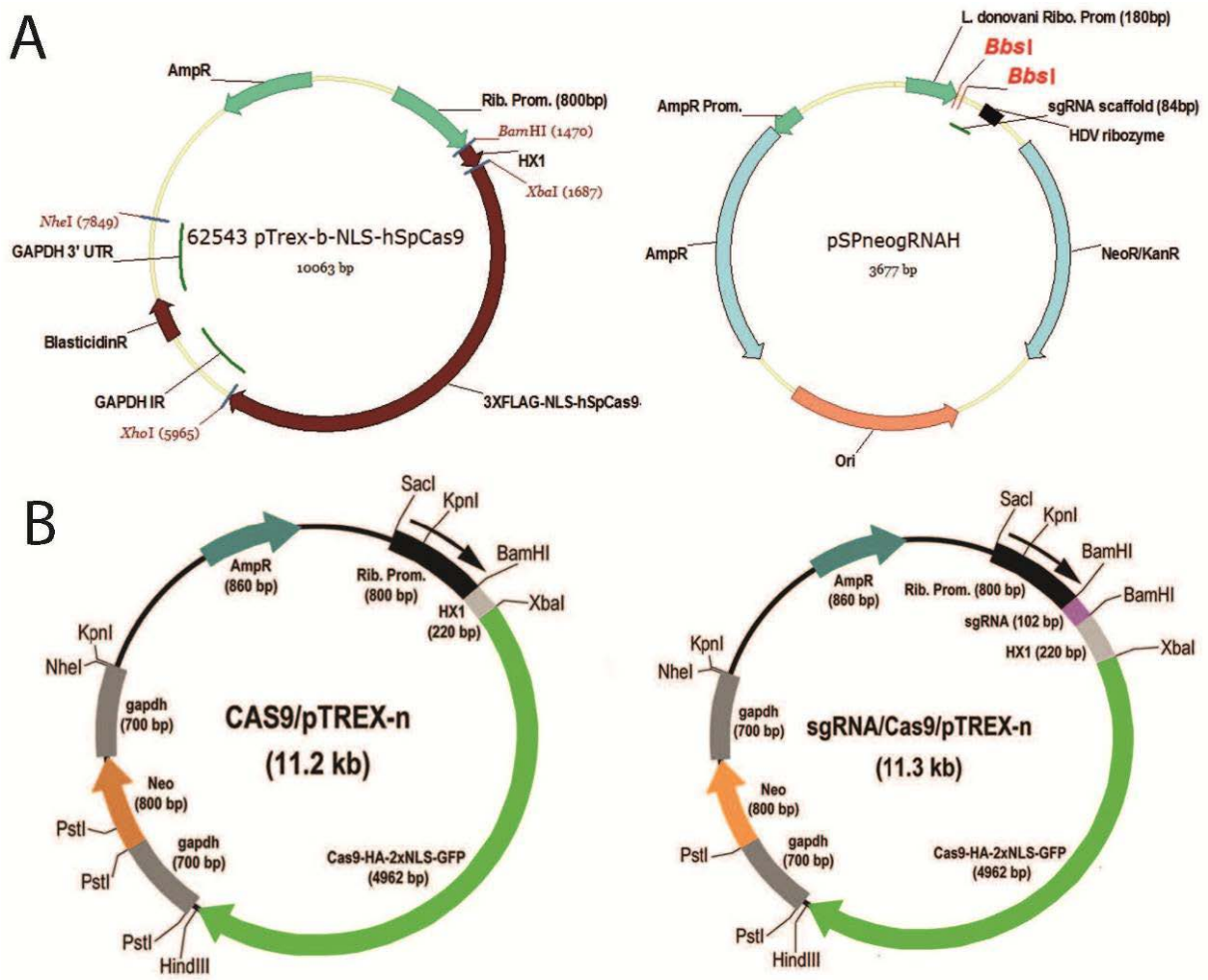


Figure 4.1 Restriction maps of molecular constructs generated for the CRISPR/Cas9 system in *L. passim*. (A) CRISPR-2V (two vectors) strategy. A specific sgRNA and the fusion gene 3XFLAG-NLS-hSpCas9 were individually cloned into two different plasmids which were used to express endonuclease Cas9 and sgRNA respectively in *L. passim* promastigotes. Specific sgRNA (20-24bp) was designed with BbsI digestion site. The specific sgRNA linker was annealed and cloned into BbsI digested pSPneogRNAH plasmid. (B) CRISPR-1V (one vector) strategy. Cas9/pTREX-n plasmid was derived from pTREX-n by insertion of the Cas9-HA-2X NLS-GFP fusion gene through XbaI and HindIII restriction sites. Plasmid sgRNA/Cas9/pTREX-n was derived from Cas9/pTREX-n by insertion of a specific sgRNA fragment through the BamHI site. Specifically, 102bp of sgRNA was amplified from pUC_sgRNA plasmid (addgene #68710) using specific forward primer and common reverse primer (more details described in materials and methods). They allow the expression of a specific sgRNA together with Cas9 through a single transfection event of wild-type *L. passim* promastigotes. Ribo. Prom, ribosomal promoter; AmpR, ampicillin resistance gene; Neo(R), neomycin resistance gene; BlasticidinR, Blasticidin-S deaminase gene; HX1 and gapdh are *trans*-splicing regions described in the past paper (Vazquez and Levin, 1999).

4.2.2. Western blot to confirm Cas9 expression

10^7 Cas9 expressing viable parasites were counted using CASY® Cell Counter together with Analyzer System Model TT from Roche Inc. 50 µL of sampling buffer with protease inhibitor cocktail was added to lyse the parasites. The cell extracts were immediately maintained at 100°C for 3.5 min, then separated by 8% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). After electrophoretic separation, polyvinylidene difluoride (PVDF) membranes were prepared by soaking into 100% methanol for 5 min and transferred

to a PVDF membrane using wet transfer apparatus according to the manufacturer's protocols (Bio-Rad) at 100 v for 90 min. The membranes were blocked in 10 mL 5% bovine serum albumin (BSA) in TBST at room temperature for 30 min with gentle shaking. The blocking buffer was discarded and ANTI-FLAG primary antibody which produced in rabbit (F7425) (diluted 1:100 in 10mL blocking buffer) were applied at 4°C overnight. PVDF membranes were washed 3 times in TBST for 5, 10, 15 min respectively. After washing, membranes were incubated with 5 mL 5% skimmed milk conjugated polyclonal IRDye® 680RD Goat anti-Rabbit secondary antibody (1:10,000) in TBST at room temperature for 1 h. The membranes were washed as mentioned above, scanned by Odyssey® scanner and analyzed.

4.2.3. sgRNA sequences design

In total, 4 genes were used as targets for CRISPR/Cas9 system. Two exogenous targets, namely, Cas9 ORF from pTrex-b-NLS-hSpCas9 and tdTomato ORF from pTrex-Neo-tdTomato. Two endogenous targets, namely *L. passim* miltefosine transporter (LpMT) and *L. passim* tyrosine aminotransferase (LpTAT). *LpMT* was selected due to this gene was well-studied in *L. donovani* CRISPR/Cas9 system (Zhang and Matlashewski, 2015;Zhang et al., 2017). *LpTAT* was selected due to our RNA-seq result since this gene was up-regulated in the early stage of infection. I may expect *LpTAT* might play a critical role for infection.

Specifically, sgRNA targeting to Cas9 ORF sequence was directly used from previous research (Peng et al., 2015) and the sgRNA linker was phosphorylated first and cloned into BbsI digested pSPneogRNAH (described later).

As for *LpMT* gene, previous research indicated that the single point mutation can result in resistance of miltefosine (Zhang and Matlashewski, 2015). The same conserved amino acid methionine (M) was also found among *L. passim* genome, in this case, the sgRNA sequence which targets to this methionine was modified in order to be complementary to the *L. passim* genome sequence.

As for *LpTAT*, search for the DNA sequence from the protein coding sequence of *L. passim* which generated by Companion pipeline. The complete sgRNA targeting sequences were designed using a custom sgRNA online design tool (grna.ctegd.uga.edu) which can identify the desired sequences and minimize the off-target effects as described previously (Peng et al., 2015). Specifically, the latest annotation gff3 file which generated by Companion pipeline

was uploaded to the website in order to predict the off-target effects with functional annotations. Specifically, as for sgRNA design, the online website 1) identifies all potential 20-bp sequences containing an NGG PAM site within the *LpTAT* sequence, 2) predict all potential off-target sites for each *LpTAT* sgRNA and those sgRNA which have 3bp or less mismatch among *LpTAT* gene. 3) predicts targeting efficiency by using a position-specific nucleotide composition scoring matrix as described before (Doench et al., 2014). A list of sgRNA sequences targeting to *LpTAT* gene was returned and the sgRNA sequences were only selected with no predicted off-target effect (i.e targeting *LpTAT* gene only) and selected by predicted efficiency score.

As for tdTomato, the complete coding sequence was used as a query for protein domain search on NCBI (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer and Bryant, 2004;Marchler-Bauer et al., 2010;Marchler-Bauer et al., 2014;Marchler-Bauer et al., 2016). Potential knock out of tdTomato may decrease the expression of red fluorescence which may detected by FACS. The gene domain tdTomato was from GFP superfamily and the domain of the sequence was used as a query for sgRNA design by using the same online sgRNA design website above.

4.2.4. Cloning sgRNA sequence into various gRNA expression vectors and *in vitro* synthesis

As for pSPneogRNAH expression vector, the complementary oligonucleotides of sgRNA sequence with BbsI digestion sites were first phosphorylated in T4 polynucleotide kinase buffer ligation buffer with T4 polynucleotide kinase and then incubated at 37 °C for 60 min, 95°C for 5 min, cooling down to 25 °C with the ramp rate at 5% in thermocycler. The annealed sgRNA linkers were then cloned into the BbsI digested pSPneogRNAH vector by using Solution I from Takara at 16 °C for 45 min. Use 6 µl ligation product to transform chemically competent TOP10 *Escherichia coli* cells following standard heat-shock transformation protocol, spread cells on LB-agar plates supplemented with 100 µg/mL ampicillin and incubated overnight at 37 °C. After transformation, analyze colonies by PCR using the specific designed forward sgRNA primer and common reverse primer. The correct amplicon size of insertion is 478 bp. After that, the specific insertion was sequenced by using pSPneogRNAH sequencing primer to verify the correct sgRNA sequence for further

transfection.

As for Cas9/pTREX-n vector, the method was followed and modified from previous research (Lander et al., 2015b). Firstly, specify the protospacer regions to the target genes and inserting the protospacer sequence $GN_{(19-24)}$ into the following primer backbone: 5'-GATCGGATCC-$GN_{(19-24)}$-GTTTTAGAGCTAGAAATAGC-3'. The common reverse primer sequence is 5'-CAGTGGATCCAAAAAAGCACCGACTCGGTG-3' which was used to amplify sgRNA. sgRNA sequence was amplified by using pUC_sgRNA (plasmid #68710) as template together with the primer set by using KOD Fx enzyme with the following condition: 1) 94 °C for 2 min, 2) 98 °C 15 s, 3) 52 °C 20 s, 4) 68 °C 20 s, 5) repeat steps 2-4 for a total of 35 cycles, 6) 68 °C for 5 min, 6) hold temperature at 4 °C. After visualized on the gel, the PCR product was purified by commercial gel extraction kits. Then, the sgRNA (122-127 bp) was cloned into BamHI digested Cas9/pTREX-n vectors by using the same ligation and transformation methods above. When performing the colony PCR, the HX1 reverse primer (5'-TAATTTCGCTTTCGTGCGTG-3') and the specific forward primer were used to detect the positive clones with sgRNA inserted in the right orientation amplify a band of 190-195 bp. Finally, the insertion was sequenced by using HX1 primer. These clones correspond to sgRNA/Cas9/pTREX-n.

The tdTomato sgRNA linker was ligated into BsaI digested pUC57-sgRNA expression vector. The recombinant colonies were randomly selected by kanamycin on agar plates for sequencing. After sequencing, plasmids with the correct sgRNA linker sequences were linearized by DraI and followed by using TranscriptAid T7 High Yield Transcription Kit (Thermo Fisher, #K0441) according to the manufacturer's instruction to synthesize sgRNA *in vitro*.

4.2.5. Overlapping Extension PCR (OE PCR) for homologous cassette containing different selective markers

For each target gene, 3 primer sets were used for OE PCR. Also, if needed, the outer primer set was used to amplify a large region across the target region and purified by commercial gel extraction kits.

4.2.5.1. *LpMT* homologous cassette containing hygromycin marker

For *LpMT* gene, *L. passim* genomic DNA was used as a template for *LpMT* outer primers

which first to amplify a large region of *LpMT* gene, which is 3,636 bp. Then, this 3,636 bp amplicon was purified by commercial gel extraction kit and used as a template for further OE PCR. Specifically, the 500 bp 5' UTR region of *LpMT* gene was selected for PCR. 500 bp of the *LpMT* ORF region which is downstream of the Cas9 cleavage site was selected for OE PCR primer design. The ORF of the hygromycin B phosphotransferase gene was also amplified by PCR from pCsV1300 vector template using primers containing minimum 9 nucleotides (nt) overlapping with the 5' and the downstream of the cleavage site of *LpMT* gene. These three fragments (5' *LpMT* flanking sequence, hygromycin-resistant gene, and downstream of *LpMT* cleavage site) were linked together by the 5' primer and 3' primer at the very end of these three PCR fragments. The final PCR product was precipitated by ethanol and then phosphorylated in T4 polynucleotide kinase as shown above. After that, the PCR product was cloned into EcoRV digested pBluescript II SK(+) (treated by Alkaline Phosphatase before ligation). After transformation, the potential clones with correct insertion were screened by blue-white screening on the agar plate. After that, analyze the white colonies by using T3 and T7 primer set. Colonies with correct insertion and orientation were digested by HindIII-HF (NEB) and precipitated by ethanol. Finally, 25 µg of the linearized drug resistance cassette was used for further co-transfection.

4.2.5.2. *LpTAT* homologous cassette containing hygromycin marker

Similarly, *L. passim* genomic DNA was used as a template for *LpTAT* outer primers which first amplify a large region flanking of *LpTAT* gene. Then the outer region, which is 2249 bp was used as a template for further PCR. Specifically, the 5' and 3' UTRs of the *LpTAT* gene (438 bp and 500 bp, respectively) were amplified by PCR using *LpTAT* outer region as a template. Similarly, the hygromycin-resistant gene was amplified as mentioned above. Three PCR fragments were ligated, cloned, sequenced and digested for further co-transfection as above.

4.2.5.3. Cas9 ORF homologous cassette containing puromycin marker

The 5' and the downstream regions of the Cas9 cleavage site (500 bp) were amplified by PCR using pTrex-b-NLS-hSpCas9 as a template. The ORF of the Puromycin N-acetyltransferase was amplified by PCR from pSpCas9n(BB)-2A-Puro (PX462) V2.0 vector as a template using primers containing at least 9 nt overlapping with the 5' and the downstream flanking sequence of the Cas9 ORF. The same method was used to obtain 25 µg of the linearized drug resistance cassette for co-transfection.

After establishing these plasmids and DNA constructs, there are three ways I could try to establish/verify CRISPR/Cas9 in *L. passim*. 1) using sgRNA synthesized *in vitro* to electroporating into Cas9 stable expression cell line (tdTomato ORF gene). 2) using one vector which can express sgRNA and Cas9 protein simultaneously (sgRNA/Cas9/pTREX-n). 3) using two vectors which express Cas9 and sgRNA on different vectors. All these methods were tested in order to establish CRISPR/Cas9 in *L. passim*.

4.2.5.4. Transfection and single clone selection for different transfectants

The parasites for electroporation were cultured in 25-cm^2 cell culture flasks (Corning Inc.) with 10 mL modified FPFB medium at 25 °C. 2×10^7 viable parasites were counted by diluted 1:500 in CASYton using CASY® Cell Counter together with Analyzer System Model TT from OMNI Life Science. Mid-logarithmic phase of *L. passim* was transfected by electroporation using GenePulserX cell electroporator (Bio-Rad). The parasites were harvested by centrifugation at 2,000 x g for 5 min at room temperature, washed with different kinds of electroporation buffers twice and resuspended in 0.4 mL the same electroporation buffer. Cells were then transferred to a 2 mm gap cuvette (Bio-Rad, USA) and 10 µg of plasmid DNA or 25 µg of sgRNA was added. As for co-transfection, 10 µg of the plasmid DNA and 25 µg of linearized cassette were used for co-transfection. The cuvettes were chilled on ice for 15 min before electroporation. There is a 10-sec interval between each pulse. The resistance was infinity to all electroporation protocols.

Three different protocols were modified and applied for electroporation from previous papers (Ngo et al., 1998;LaCount et al., 2002;Alonso et al., 2014). Different voltage gradients with different electroporation buffers combinations were used to improve the electroporation efficiency in *L. passim*. Specifically, the components of the low electroporation buffer are as follow: 140 mM NaCl, 25 mM HEPES and 0.74 mM Na_2HPO_4, pH 7.5. The electroporation parameters are as follows: 2 pulses at 400 V and 500 µF were applied to the cuvettes as described above. As for high voltage protocol, there are 2 protocols for this. Specially, the Cytomix contains: 120 mM KCl, 0.15 mM $CaCl_2$, 10 mM K_2HPO_4, 25 mM HEPES and 5 mM $MgCl_2$, pH 7.6. As for EM buffer, it is a 3:1 mixture of Cytomix and phosphate-sucrose buffer (277 mM sucrose, 1mM $MgCl_2$, 7 mM K_2HPO_4, pH 7.4). KH_2PO_4 was used to adjust the pH level for the phosphate-sucrose buffer. All these electroporation buffers were filtered through a 0.22 µm filter and stored at 4°C.

pTrex-Neo-tdTomato, pTrex-n-eGFP, and pTrex-b-NLS-hSpCas9 were electroporated into *L. passim* in order to obtain tdTomato, GFP and Cas9 (Blasticidin resistance) expression cell lines according to the low voltage transfection protocol. Plasmids pLPhygCAS9 and pSPneogRNAH were electroporated into *L. passim* in order to obtain Cas9 (Hygromycin resistance) expression and potential MMEJ KO or chromosomal translocation cell lines by using high electroporation buffer. Generally, there are 2 strategies to establish CRISPR/Cas9 system in *L. passim*.

Firstly, *L. passim* which can stable express Cas9 protein can be electroporated with sgRNA which synthesized *in vitro* targeting to the desired sequences on the genome.

Secondly, stable Cas9 expressing *L. passim* could be transfected with plasmids which contain sgRNA cloning site (2 vector strategy) or, the sgRNA cloning site together with Cas9 ORF on the same plasmid (1 vector strategy). In order to establish CRISPR/Cas9 system in *L. passim* using different strategies and verify the validity of CRISPR/Cas9 system in *L. passim*, the following plasmids combinations were tested for CRISPR/Cas9 system in *L. passim*:

For the two-vector strategy, the stable Cas9-expressing cell line was co-transfected with pSPneogRNAH which may result in the DSB near the cleavage sites and repaired by MMEJ pathway or chromosomal translocation. Later, the drug resistance cassette was co-transfected together with sgRNA expression vector which induces HDR caused by CRISPR/Cas9 in *L. passim*.

For the one-vector strategy, sgRNA/Cas9/pTrex-n was directly electroporated into *L. passim* WT which can express Cas9 protein and sgRNA simultaneously.

After electroporation, the cuvettes were maintained on ice until being transferred into 5mL modified FPFB medium in a T25 flask. Antibiotics were added to the medium for selection after 24 h. Specifically, 5 μg/mL of blasticidin, 400 μg/mL of puromycin, 1 μg/mL of hygromycin and 50 μg/mL of G418 were added to the medium according to different selection purposes. The electroporated parasites were centrifuged and resuspended with new medium every 3 days and the concentration of antibiotics was maintained the same for selection. The drug concentration was maintained for 4 weeks. The selection of recombinant colonies was conducted as follows: 5% of agarose was sterilised by autoclave and cooled down to 70 °C. The modified FPFB medium was heated to 50 °C. An equal volume of the sterilised agarose and the medium were well-mixed immediately with the concentration of antibiotics as mentioned above. The mixture was instantly transferred into petri dish until

solidified. The diluted parasites from the medium were spread on the sterilised agarose plate and incubated at 25 °C. The recombinant parasite colonies were picked and transferred to 12-well plate with the modified FPFB medium containing the proper antibiotics concentration which mentioned above. Then, isolate gDNA from transfectants and analyze them individually by PCR to verify gene disruption. If heterozygous are detected, transfectants should be maintained for 2 more weeks with double the concentration of antibiotics and then re-analyzed by PCR to confirm the gene disruption in the entire population. The primer sets which used for analyzing the gene disruption can be found in the results and the specific sequences can be found in the Appendix 1.

For KO with fluorescence marker, before flow cytometric analysis, *L. passim* wild type and the stable tdTomato cell line were analyzed by flow cytometer first to determine the positions of these two fluorescence peaks. The gate was used to select the viable cell according to the proper morphology. Previous research demonstrates that using the high voltage method can introduce more exogenous DNA compared with the low voltage method in *Leishmania* species (Robinson and Beverley, 2003). In order to optimize the best electroporation condition for *L. passim*, pTrex-Neo-tdTomato was used as a fluorescence marker to increase electroporation efficiency. Both EM buffer and Cytomix were used to introduce plasmids into the parasites. However, only Cytomix was used for flow cytometric analysis. Different voltage gradients were applied at 600 V, 900 V, 1,200 V, 1,500 V and 1,800 V. After 48 h, 72 h, 96 h and 192 h, parasites were diluted in PBS followed by flow cytometric analysis. 30,000 cells were analyzed by flow cytometer. All the experiments were performed in triplicate.

4.2.6. Flow cytometry and fluorescence microscopy

Flow cytometric analysis was performed using a FACSCalibur™ flow cytometer from Becton Dickinson (BD Biosciences, USA). Data were collected by using BD FACStation™ software.

Fluorescence microscopy was performed to determine the presence of tdTomato (red fluorescence). Images were acquired with NIKON ECLIPSE Ni-U and Ti-S microscope. Images were adjusted for contrast using Adobe Photoshop CC and Adobe Illustrator CC. The tdTomato stable expression cell line was cultivated in two flasks with or without the existence of G418. If the concentration reached the stationary phase, old medium was removed and the new medium was added as well as the antibiotics. The cells were

maintained regularly for photographing by NIKON ECLIPSE Ni-U or Ti-S microscope.

4.2.7. T7 endonuclease I (T7EI) analysis for *LpTAT* and *LpMT* genes using sgRNA expression vector only

I also investigated if CRISPR/Cas9 system I established in *L. passim* could result in indel near the cleavage sites for *LpTAT* and *LpMT* genes. In order to harvest more transfectants with sgRNA expression vector only, after electroporation, doubled EC50 of each antibiotic was added to the medium in order to fast eliminate any WT parasites. New medium was replaced together with the antibiotics every four days. This process is to enhance the expression of sgRNA expression vector in the transfected mutants. When the concentrations became higher, pooled parasites were directly used for DNA extraction. After extracting the DNA, 500 bp and 1000 bp near the cleavage sites for *LpMT* and *LpTAT* genes were amplified for together with WT. PCR products with the expected sizes were purified by commercial gel extraction kit. T7EI was bought from NEB and the PCR products were treated as the manufacturer's instruction on a bio-rad thermo cycler.

4.2.8. mRNA detection for *LpMT* and *LpTAT* homozygous and heterozygous KO

After obtaining the *LpMT* and *LpTAT* homozygous and heterozygous KO stains, total mRNA was extracted by TRIzol reagent. The concentrations of total RNA were measured by nano-drop. All the concentrations of total RNAs were diluted to around 200 ng for downstream analysis. Before reverse transcription, total RNAs from each strain were treated by 1 U of RNase-free DNase from Promega in 10 µL at 37 °C for 30 min. Then, 1 µL of the stop reaction was added to terminate the reaction. This process was used to eliminate the potential gDNA contamination during total RNA extraction. 20 µL DEPC H_2O was added to each reaction for reverse transcription. After establishing the cDNA library for each strain, gene-specific primer sets were used to confirm the mRNA expression. *L. passim* GAPDH was used as a positive control (Kojima et al., 2011).

4.2.9. Comparing the KO efficiency of homologous recombination and CRISPR/Cas9 system in *L. passim* targeting *LpMT* gene

In order to compare the efficiency of homologous recombination and CRISPR/Cas9 system in *L. passim*, I used *LpMT* gene as a target to compare the homozygous KO rate for both methods. Specifically, for the homologous recombination, 25 µg of the linearized homologous cassette was electroporated into 2 x 10^7 *L. passim* WT cells. As for the CRISPR/Cas9 method, 10 µg of sgRNA expression vector together with 25 µg linearized homologous cassette were used as described above. The same electroporation condition was used and the parasites were selected for 2 months in order to obtain the homozygous KO from these cell populations. The medium was replaced regularly and the same concentration of antibiotics was added to the medium when replaced the medium. After selection, these transfectants were spread in the 96-well plate in order to obtain the single clone. Then, these single clones were analyzed by PCR to verify the KO on the genome.

4.3. Results

4.3.1. The growth of *L. passim* in FPFB medium

In order to determine the logarithmic phase (highly active) of *L. passim* for both transfection and infection, the growth curves of both *L. passim* WT and Cas9 stable expressing cell lines were measured which can be found in figure 4.2. The growth curves of both strains are identical to each other which indicate the expression of Cas9 does not suppress the growth of *L. passim in vitro*.

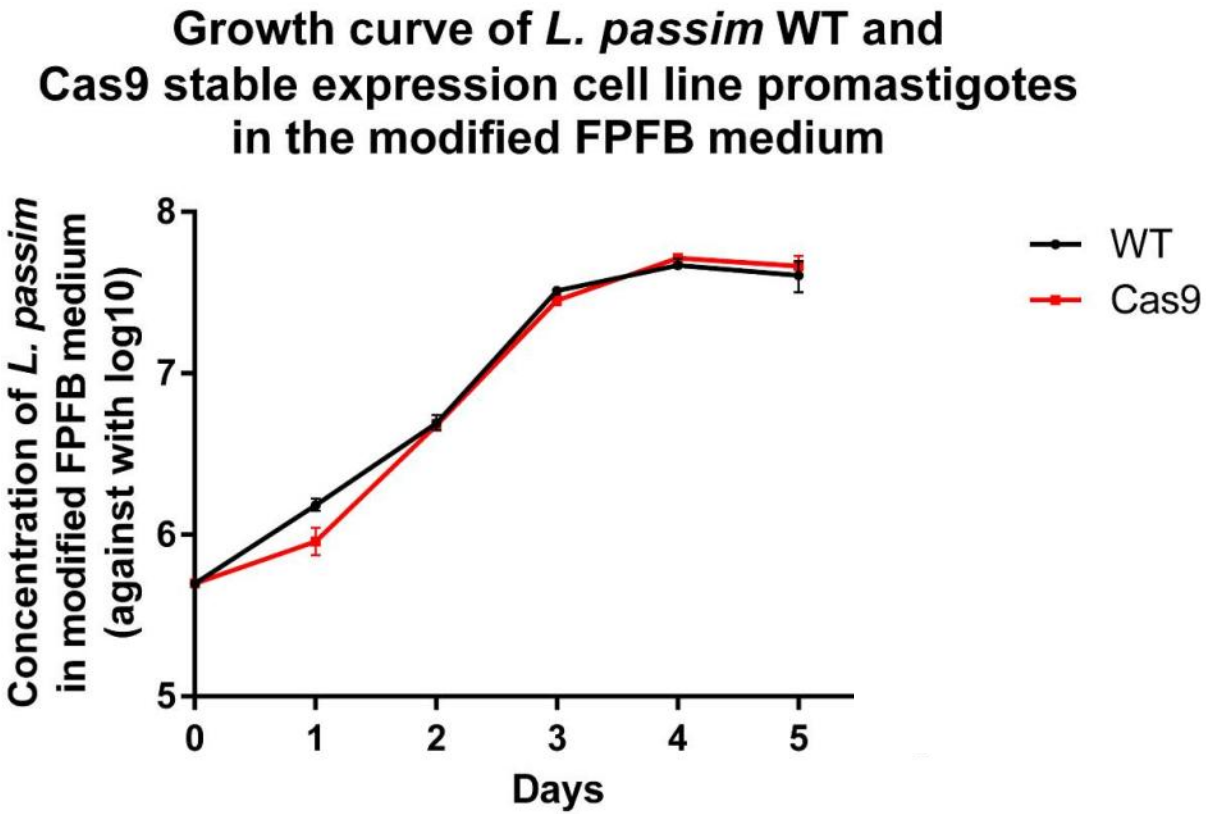


Figure 4.2.The growth curves of *L. passim* WT and Cas9 stable expression promastigotes in modified FPFB medium. The experiment was performed in triplicate. The initial concentration of *L. passim* was 5×10^5/ml. It takes 6 days for *L. passim* to grow to the stationary phase. The concentration of *L. passim* in the stationary phase was about 6×10^7/ml. From the growth curve, it is obvious that the logarithmic phase of *L. passim* growth is at 48-72 h after the culture. The error bars indicate the standard deviations.

4.3.2. Expression of fluorescent proteins in *L. passim* and the single clone selection on agarose plate

After electroporation with pTrex-Neo-tdTomato and pTrex-n-eGFP, single colony was isolated on agarose plate with antibiotics. After spreading on agarose for 15 days, Figures 4.3 and 4.4 indicate the electroporation protocol can be used to introduce plasmid DNA into *L. passim*. Both tdTomato and GFP can be used as fluorescence markers in *L. passim*. The one with the highest red fluorescence was selected for the further experiment.

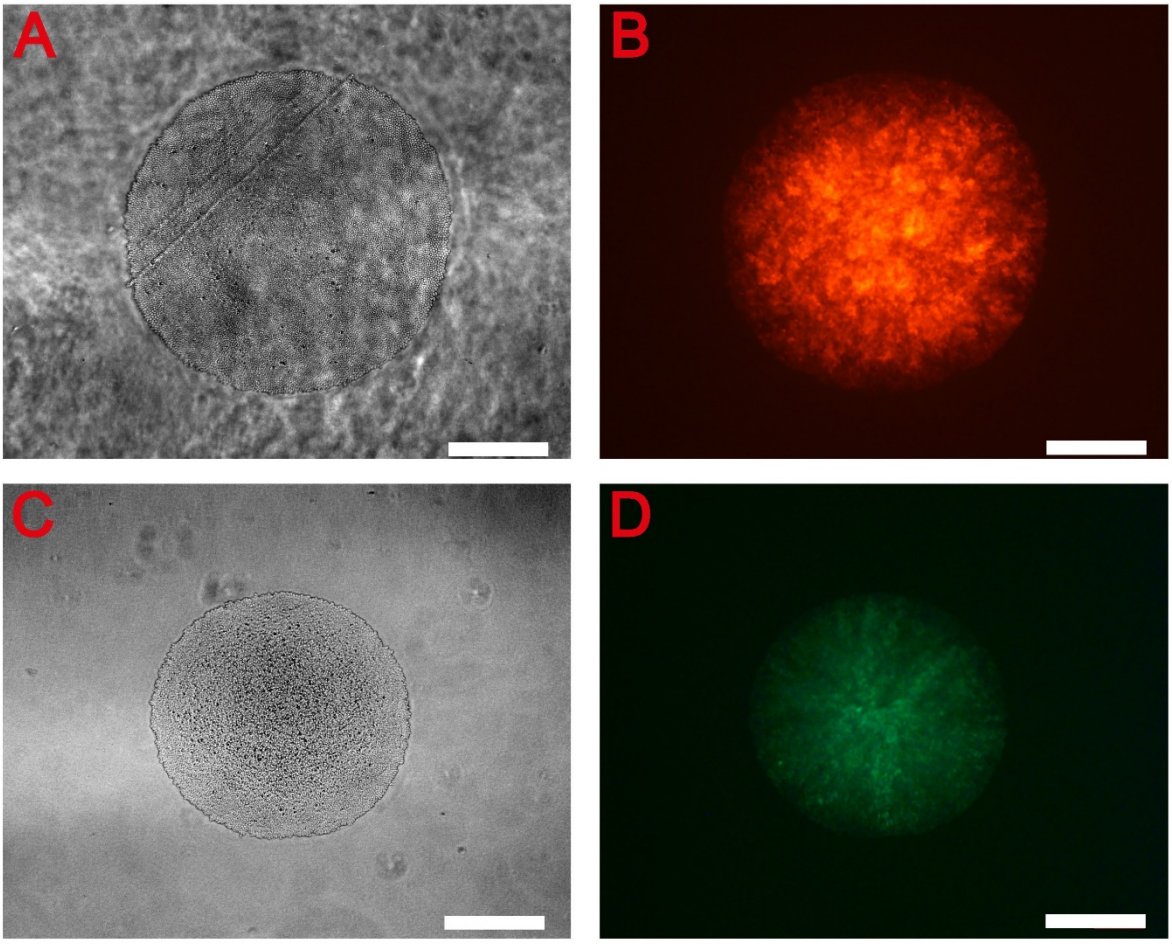


Figure 4.3 Both tdTomato and GFP expressing parasite lines can grow on the agarose plate. It takes 10 to 12 days until the colony becomes visible by eye. (A) Td-Tomato expressing clone on agarose plate (Phase). Adjusted to greyscale by Photoshop 2017 CC. (B) Td-Tomato expressing clone on agarose plate (Fluorescence). (C) GFP expressing clone on agarose plate (Phase). Adjusted to greyscale by Photoshop 2017 CC (D) GFP expressing clone on agarose plate (Fluorescence). Also, in *L. passim*, the fluorescence intensity of tdTomato is higher than that of GFP. Inverted microscopy. 40X. Bar = 200 µm.

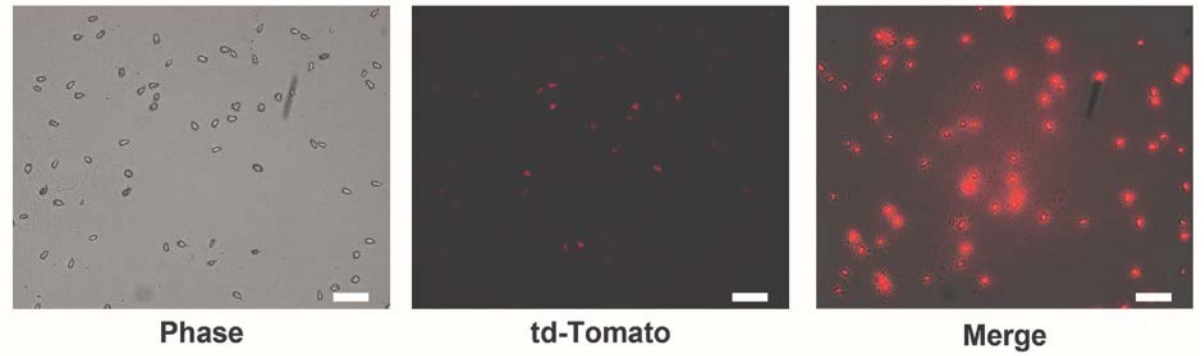


Figure 4.4 tdTomato expressing parasite cells from the single colony on agarose plate. All cells have red fluorescence indicating single clone selection was successful. Inverted microscopy. 400X. Bar=30 µm.

4.3.3. Improvement of transfection efficiency of *L. passim*

From Figure 4.5, it is possible to separate tdTomato-expressing parasites from the wild type and test the transfection efficiency by electroporation.

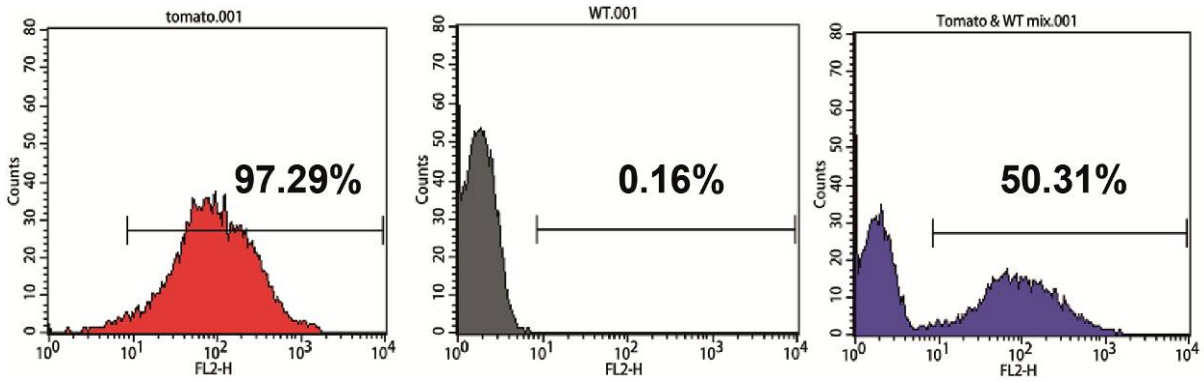


Figure 4.5 Flow cytometric analysis of tdTomato expressing parasite line, wild type (WT), and a 1:1 mixture of WT and tdTomato. As seen in Tomato & WT mix, the tdTomato expressing cells can be separated with the WT which can be used as a fluorescence marker. 10,000 cells were counted.

Figures 4.6, 4.7, 4.8 and 4.9 indicate the transfection efficiency of *L. passim* by electroporation. Specifically, 48h after electroporation, tdTomato-positive parasites cannot be seen under the microscope. From the flow cytometric analysis, the transfection efficiency was less than 0.21% under all conditions, which may reveal as for *L. passim*, it takes more than 48 h until the protein could be synthesized. Interestingly, 72 h after electroporation, tdTomato-positive parasites could be observed under the microscope. 96h later, the highest efficiency is 1.21% at the voltage of 1.8 kv. 192 h after electroporation, although red fluorescence can be detected by flow cytometer, the percentage of fluorescent cells decreased from all groups. This indicates that *L. passim* tends to lose the expression of tdTomato after the transfection. The same phenomenon has also been reported in *L. donovani* (Robinson and Beverley, 2003). The groups applied with the highest voltage have the most fluorescent cells.

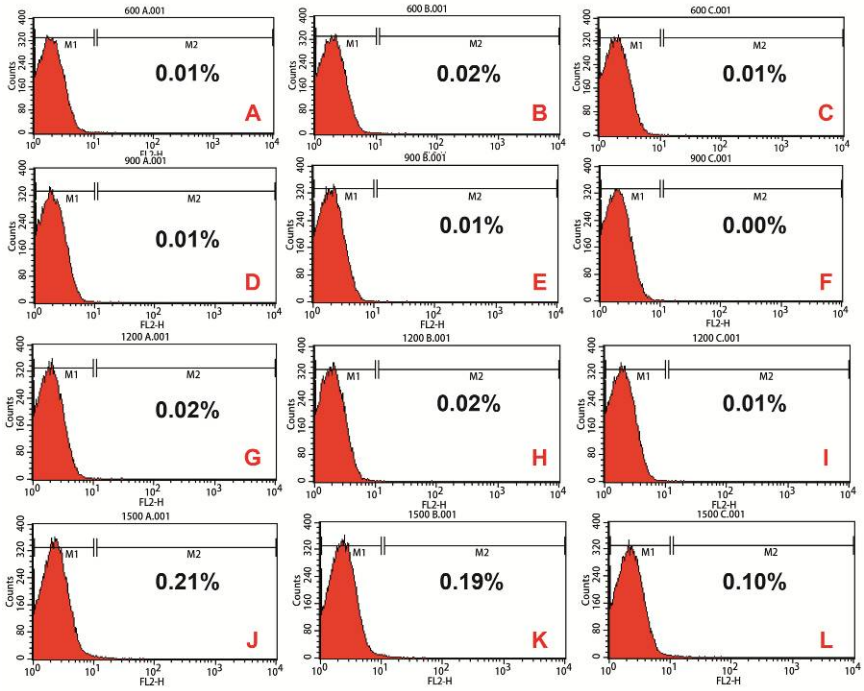


Figure 4.6 Transfection efficiency of *L. passim* electroporated with different voltages using Cytomix buffer at 48 h after electroporation. Each experiment was conducted in triplicates. A, B and C: 600 v; D, E and F: 900 v; G, H and I: 1.2 kv. J, K and L: 1.5 kv. 30,000 cells were counted.

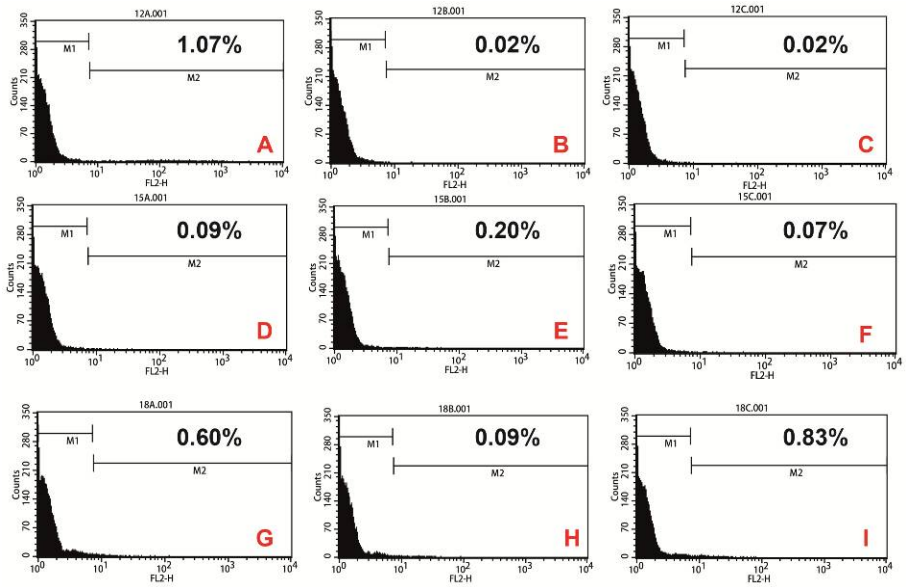


Figure 4.7 Transfection efficiency of *L. passim* electroporated with different voltages using Cytomix buffer at 72 h after electroporation. Each experiment was conducted in triplicates. A, B and C: 1.2 kv; D, E and F: 1.5 kv; G, H and I: 1.8 kv. 30,000 cells were counted.

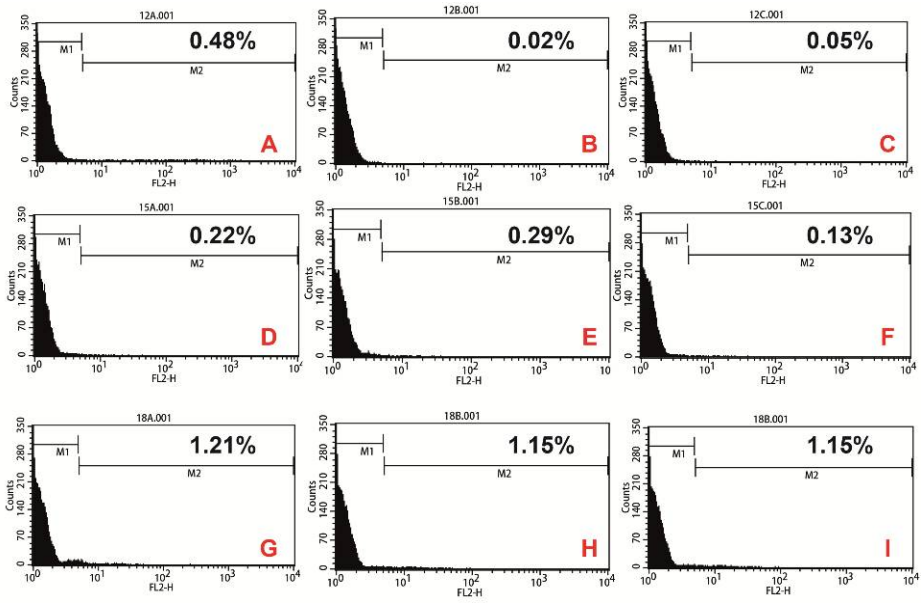


Figure 4.8 Transfection efficiency of *L. passim* electroporated with different voltages using Cytomix buffer at 96 h after electroporation. Each experiment was conducted in triplicates. A, B and C: 1.2 kv; D, E and F: 1.5 kv; G, H and I: 1.8 kv. 30,000 cells were counted.

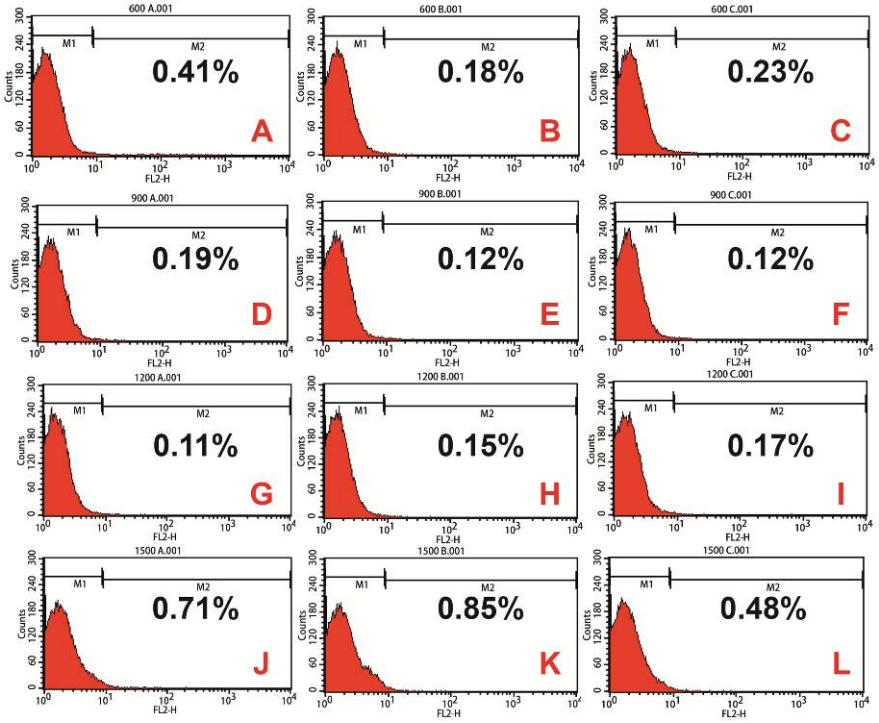


Figure 4.9 Transfection efficiency of *L. passim* electroporated with different voltages using Cytomix buffer, 192 h after electroporation. Each experiment was conducted in triplicates. A, B and C: 600 v; D, E and F: 900 v; G, H and I: 1.2 kv. J, K and L: 1.5 kv.30,000 cells were counted.

From the results above (Figure 4.6, 4.7, 4.8 and 4.9), the high voltage is better to introduce exogenous plasmid DNA into *L. passim* compared with low voltage. This might be due to the high voltage generates larger pores on the cell membrane compared with low voltage.

4.3.4. Transfected plasmid DNA was not integrated in *L. passim* genome

After 35 weeks' (67 passages) cultivation in the flask without G418, the majority of the transfected parasites lost the red fluorescence (Figure 4.10). This indicates pTrex-Neo-tdTomato plasmid was not integrated into the genome of *L. passim*. Moreover, this phenomenon also suggests that it is necessary to add the antibiotics in the medium to maintain the transfected plasmid DNA in *L. passim*.

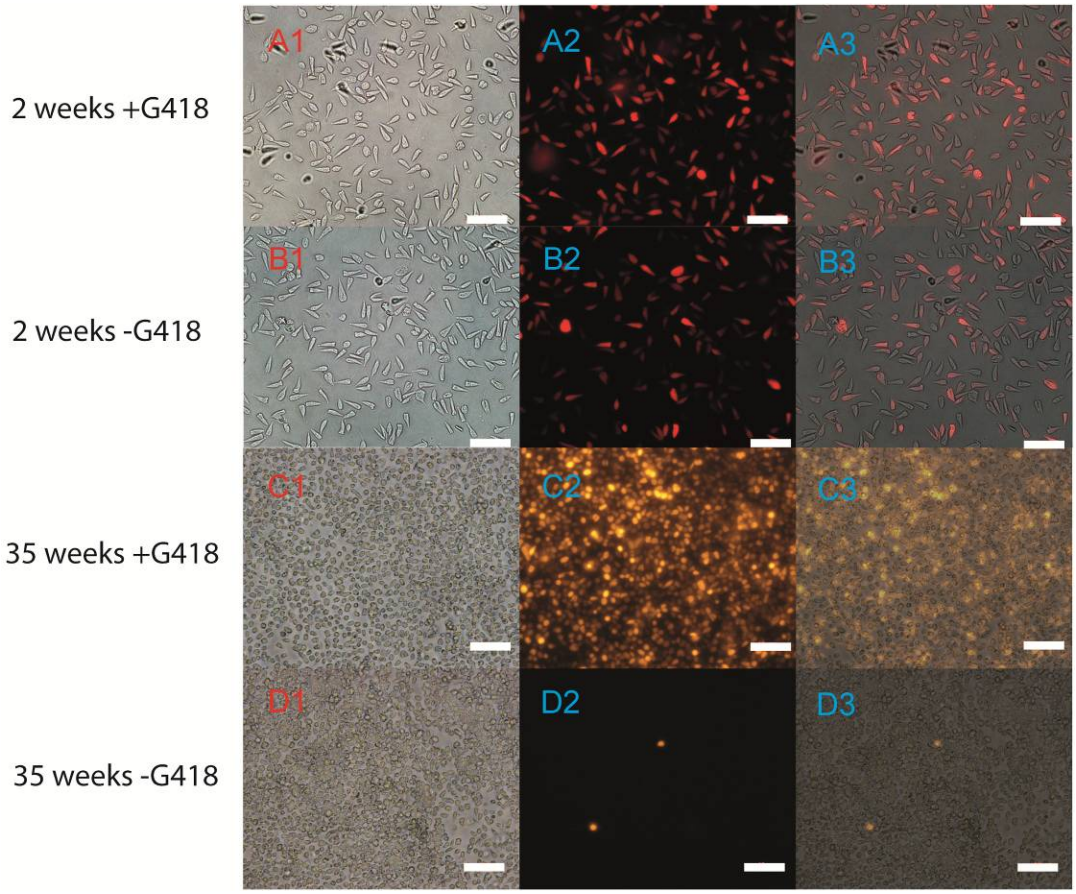


Figure 4.10 tdTomato-expressing plasmid is lost in the *L. passim* after 35 weeks (67 passages) without antibiotics selection. From left to right: phase, fluorescence, merge. Bar = 50 µm.

4.3.5. Establishment of CRISPR/Cas9 in *L. passim*

4.3.5.1. Expression of Cas9 and sgRNA in *L. passim*

To enable CRISPR/Cas9 gene editing, Cas9 endonuclease should be expressed by *L. passim* promastigote first. *L. passim* was transfected with an expression plasmid (pTrex-b-NLS-hSpCas9) carrying the humanized *Streptococcus pyogenes* Cas9 nuclease gene (hSpCas9) with a nuclear localization signal (NLS) and three copies of the FLAG epitope tag at the N-

terminus, together with a blasticidin resistance gene driven by *T. cruzi* ribosomal promoter and HX1 enhancer (Vazquez and Levin, 1999;Cong et al., 2013;Peng et al., 2015). Western blotting confirmed the expression of Cas9 in the blasticidin-resistant cell line (Figure 4.11). The number 6 clone was randomly chosen for the further experiments in order to establish the CRISPR/Cas9 system in *L. passim*.

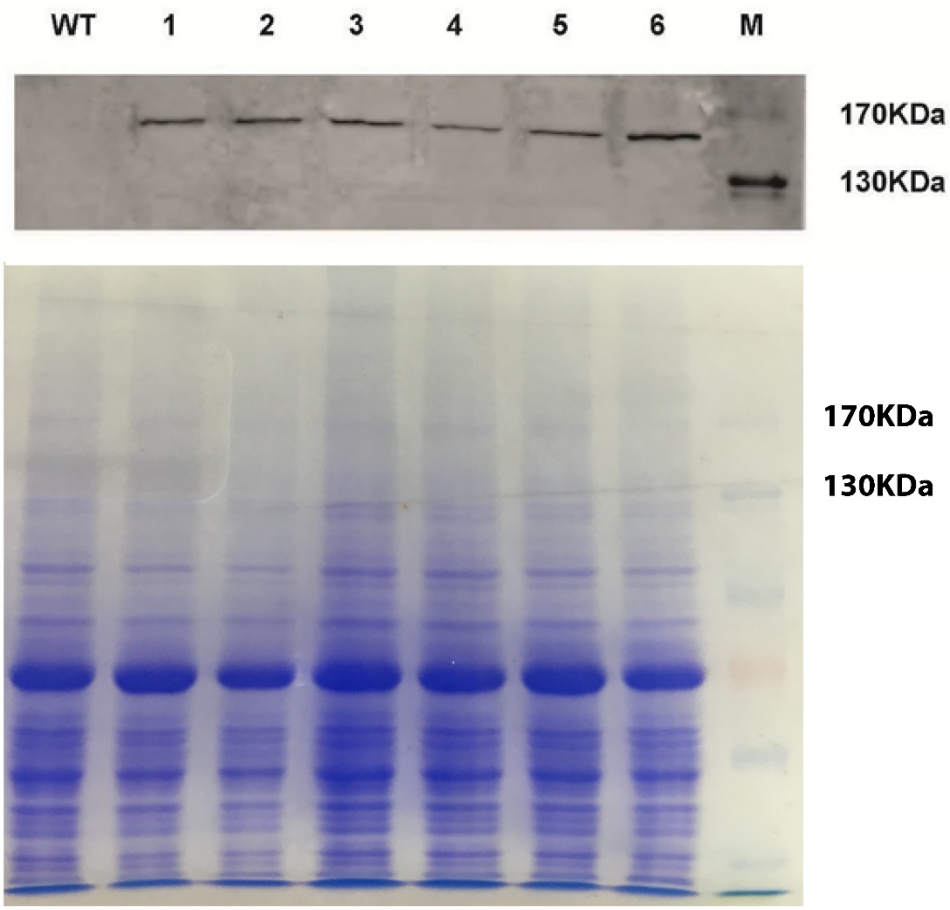


Figure 4.11 After transfection with Cas9 expressing plasmid DNA, six blasticidin resistant colonies were picked from agarose plate and expanded. 10^7 parasites were used for the experiment. WT: wild type *L. passim*; Lane 1-6: Six different clones; M: Molecular weight marker. The figure below shows the Coomassie Brilliant Blue (CBB) staining as a loading control.

4.3.5.2. *L. passim* CRISPR/Cas9 system using sgRNA expression vector only

In order to test if CRISPR/Cas9 system can be used to induce any genome editing in *L. passim*, initially, only sgRNA expression vector was applied to test if it can induce indels or chromosomal translocations on the *L. passim* genome. It has been shown that by using sgRNA only, it can induce indels and chromosomal translocations near the sgRNA cleavage site in *L. donovani* (Zhang and Matlashewski, 2015;Zhang et al., 2017).

4.3.5.3. Attempt to knock-out Cas9 and *LpMT* by sgRNA expression vector only

Since introducing sgRNA synthesized *in vitro* to *L. passim* did not work because of the low transfection efficiency, I tested if one-vector strategy can trigger the MMEJ pathway like in *L. donovani* (Zhang and Matlashewski, 2015), *T. cruzi* (Peng et al., 2015) or chromosomal translocation like in *T. brucei* (Lander et al., 2015a). CRISPR/Cas9 system was successfully used to knock-out exogenous Cas9 in *T. cruzi* (Peng et al., 2015). Figure 4.12 indicates that the 500 bp near the cleavage site for 12 single clones are intact as in the WT. I also directly sequenced these PCR products which indicate they have the same sequence as WT. I also sequenced a larger region near the cleavage site with the size of 4 kbp. Still, all the 12 clones show a 4k bp band which is identical to the Cas9 expressing parasite line.

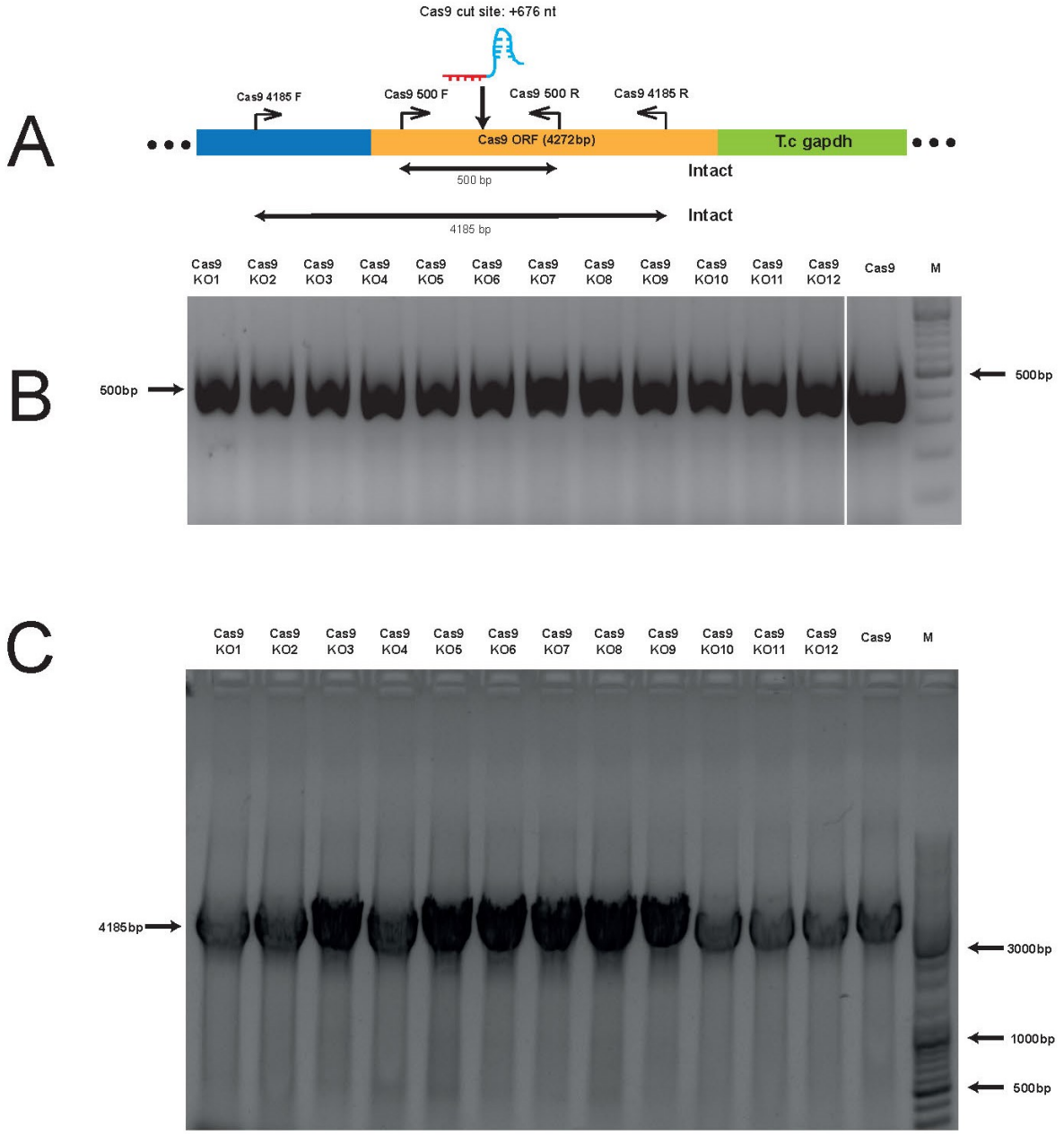


Figure 4.12 Schematic representation of the strategy to knock-out Cas9 gene by inducing the potential MMEJ pathway. (A) The positions of primers used to identify the potential deletions in Cas9 gene induced by CRISPR/Cas9. Sizes of the two amplicons are 500 bp and 4,185 bp, respectively. The putative cleavage site by Cas9 is at 676 of the 3xFLAG NLS-hSpCas9 ORF. (B) 12 individual clones picked up from the agarose plate were analyzed by PCR to generate 500 bp amplicon together with the Cas9-expressing parasites. (C) Detection of the 4,185 bp fragment by PCR. M: 10 kb molecular ladder.

4.3.5.4. *LpMT* KO with Cas9 and sgRNA expression

For the endogenous *LpMT*, figure 4.13 indicates that the 500 bp genomic DNA fragment near the putative cleavage site was intact in both isolated clones and WT. Two single colonies were selected to further investigate the potential changes in *LpMT* gene. Two large fragments which locate upstream (1575 bp) and downstream (2102 bp) of the putative

LpMT cleavage site were analyzed by PCR. However, two isolated clones have the same expected size as WT. This may further indicate using sgRNA expression vector only can trigger neither MMEJ nor chromosomal translocation on the endogenous gene of *L. passim*.

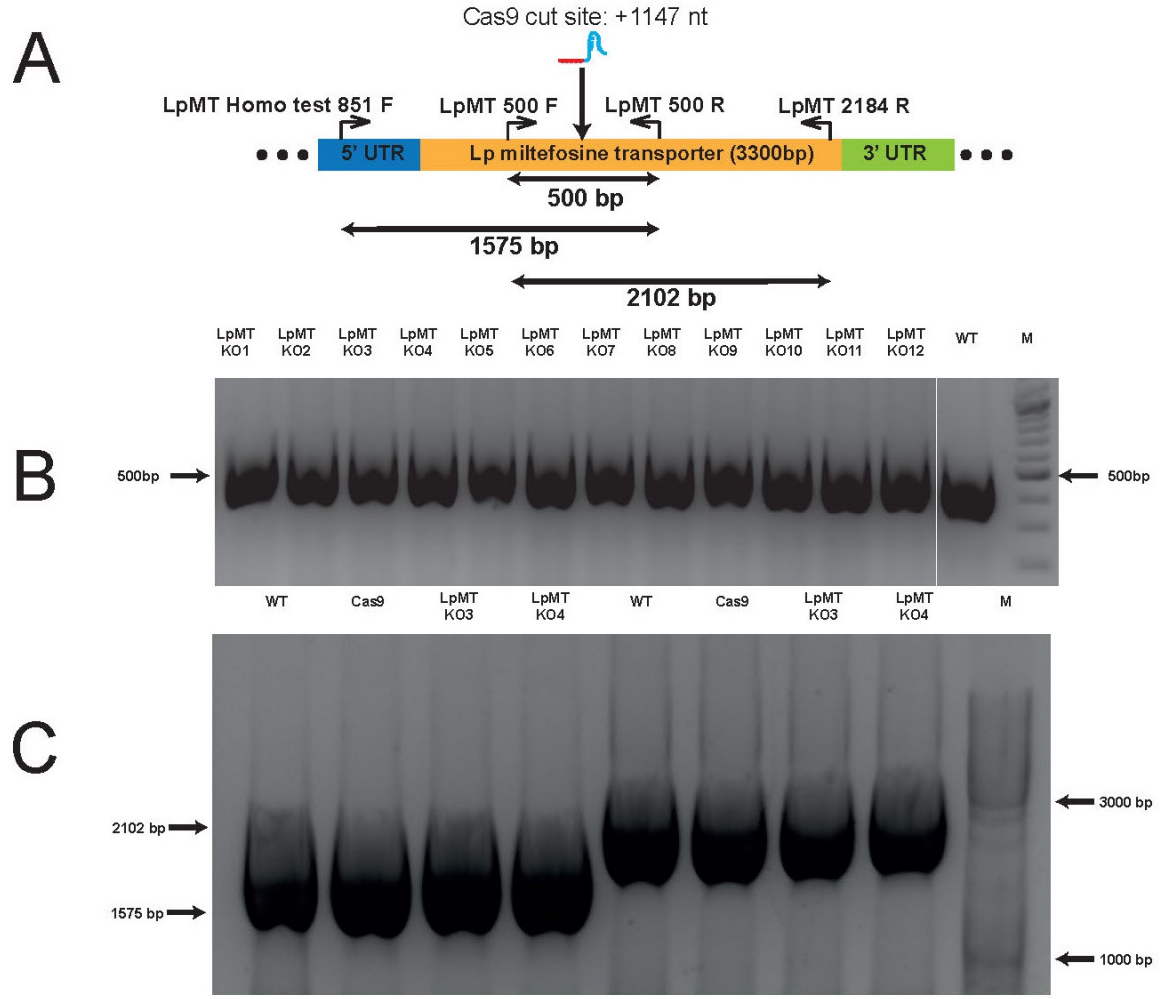


Figure 4.13 Schematic representation of the strategy used to test the potential existence of MMEJ pathway with the endogenous gene *LpMT* using pSPneogRNAH vector. (A) Primer sets used to identify the potential deletion on *L. passim* genome induced by CRISPR/Cas9 near the putative Cas9 cleavage site. The sizes of these amplicons are 500 bp, 1575 bp and 2102 bp, respectively. The sgRNA target site of Cas9 is at 1147 from the start codon of *LpMT* ORF. (B) 12 individual clones were picked up from the agarose plate for PCR using primers for 500 bp amlicon together with the WT. All these 500bp amplicons were sequenced and there were no indels near the putative cleavage sites. (C) The clones 3 and 4 were selected for analysis of the large regions for the potential chromosomal translocation together with WT and Cas9 stable expression parasite lines. Lane 1 to 4: 1575 bp amplicon. Lane 5 to 8, 2102 bp amplicon.

4.3.5.5. Testing induction of DSB in *LpMT* and *LpTAT* genes by sgRNA expression with T7E1

T7EI, which is a structure-specific nuclease that cleaves DNA hetero-duplexes at mismatch sites, I used to analyze the existence of subtle mutations induced by sgRNA expression vector only. I performed the T7EI analysis with the pooled parasites after electroporation of

sgRNA expression vector targeting to *LpTAT* and *LpMT* genes. Figure 4.14 shows that there is no difference between WT and the pooled samples after T7EI treated for both *LpMT* and *LpTAT* genes. I also tried different T7EI treatment method as in *Drosophila* (Kondo and Ueda, 2013); however, there is no change near the putative cleavage site for both endogenous genes (Not shown). These results indicate the expression of sgRNA targeting *L. passim* endogenous genes did not induce any subtle changes near the target sites.

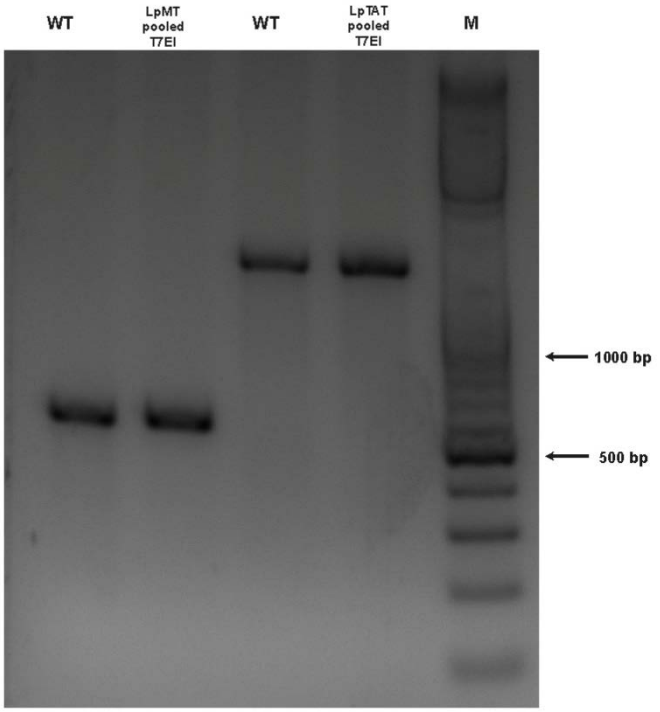


Figure 4.14 T7EI analysis of the pooled parasite samples transfected by sgRNA expression vector targeting *LpMT* and *LpTAT*. PCR products with correct size were purified from agarose gel and 200 ng of PCR products were treated with T7EI and visualized on 2% gel. Lane 1 and 2: *LpMT* 500 bp region from WT and the pooled parasites; Lane 3 and 4: *LpTAT* 1106 bp region from WT and pooled parasites; M: 100 bp molecular marker.

4.3.5.6. *L. passim* CRISPR/Cas9 system using sgRNA expression vector together with donor DNA

From above results, introducing sgRNA expression vector only did not induce any changes on the *L. passim* genome. Thus, the homologous donor might be necessary to facilitate the homologous recombination. I decided to test if gene knock-out is possible with HDR induced by CRISPR/Cas9 system.

4.3.5.7. Cas9 self-knockout by donor DNA with puromycin-containing cassette

Figure 4.15 indicates the correct insertion of the puromycin resistance gene at Cas9 gene.

However, all colonies also contain the WT Cas9 sequence. From the western blot in figure 4.15 (D), these clones still express Cas9 protein. It is likely that there are multiple copies of pTrex-b-NLS-hSpCas9 plasmids in the transfected *L. passim* and some of them were KO by CRISPR/Cas9 system whereas some of them were intact. These results indicate that with the help of homologous cassette, the two-vector strategy can be used to disrupt the *L. passim* exogenous genes.

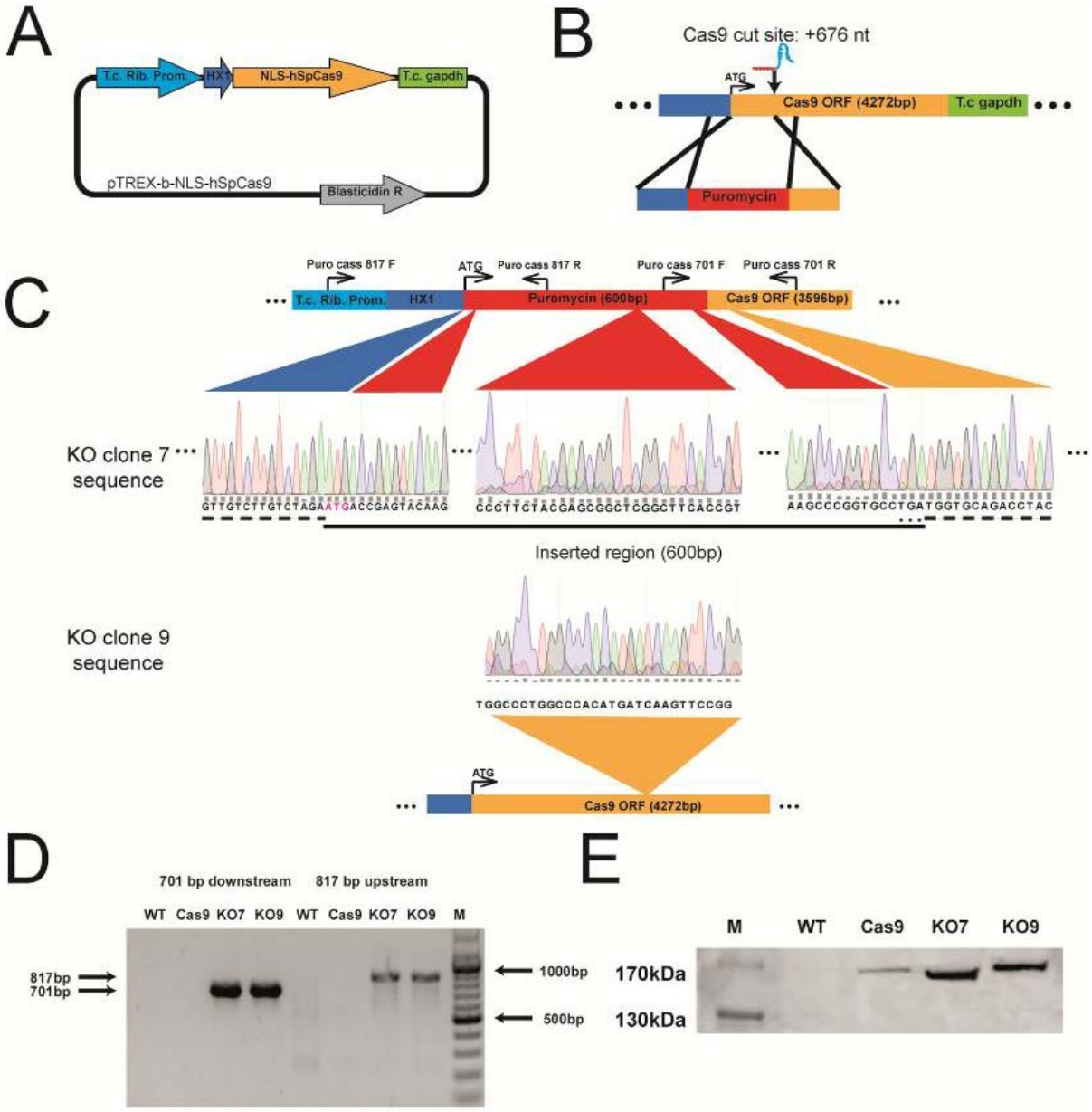


Figure 4.15 Cas9-mediated Cas9 self-disruption in *L. passim*. (A) Schematic representation of pTrex-b-NLS-hSpCas9 for the stable expression of nuclear Cas9 in *L. passim* (Addgene 62543). (B) Schematic representation of the strategy used to generate a Cas9-KO mutant by CRISPR/Cas9-induced homologous recombination. The putative cleavage site by Cas9 is at 676 of the 3xFLAG NLS-hSpCas9 ORF. The double strand break was expected to induce homologous recombination between Cas9 and the Puromycin N-acetyltransferase cassette flanking ~500 bp homologous regions from the 3xFLAG NLS-hSpCas9 ORF cleavage site. (C) Schematic representation of the recombinant Cas9 KO strain. Primers (arrows) that were used to verify gene disruption by PCR. The upstream PCR amplicon is 817 bp, while the downstream PCR amplicon is 701 bp. Electrogram showing precise insertion of the puromycin N-acetyltransferase cassette into the Cas9 cleavage

site by HDR. The upper electrogram represents the sequence of KO clone No.7, which the Cas9 ORF was disrupted by puromycin cassette. Whereas the lower electrogram represents the sequence of KO clone No.9, which indicates HDR occurred with a few copies of Cas9 ORF. Both electrograms show multiple peaks which indicate the mixture of the Cas9 KO and the Cas9 stable expression cell line. The magenta coloured ATG and the three asterisks indicate the start codon and the stop codon of puromycin N-acetyltransferase ORF respectively. (D) PCR products obtained with genomic DNA from different parasite lines using the upstream and downstream primer sets were analyzed by electrophoresis on 2% agarose gel. Lanes: WT, *L. passim* wild type; Cas9, Cas9 stable expression cell line; KO7, Cas9 KO clone No.7; KO9, Cas9 KO clone No.9; M: 100 bp to 10 kb molecular ladder. (E) Western blot analysis of protein lysates from different parasite lines. Lanes: M, protein maker; WT, *L. passim* wild type; Cas9, Cas9 stable expression cell line; KO7, Cas9 KO clone No.7; KO9, Cas9 KO clone No.9.

4.3.5.8. *L. passim* tyrosine aminotransferase (*LpTAT*) KO by CRISPR/Cas9 with hygromycin-containing cassette

Before using puromycin as a selective marker after transfection, I first measured the EC50 of puromycin for *L. passim* strain SF. Unexpectedly, this parasite has an extremely high tolerance to puromycin. The EC50 of puromycin for *L. passim* is 400 μg/mL. Under this circumstance, I decided to use hygromycin instead of puromycin. From our RNA-seq results, *LpTAT* gene was significantly upregulated at 7 days after infection in the honey bee. The previous report indicated that *LiTAT* is a novel drug target candidate for *L. infantum* (Moreno et al., 2014). I also assumed that *LpTAT* may also play critical roles for *L. passim* to establish infection and colonize in honey bee hindgut. Thus, I used CRISPR/Cas9 system together with hygromycin homologous cassette to KO this gene. Figure 4.16 shows that the entire *LpTAT* gene was replaced by hygromycin cassette. The replaced locus is around 200bp shorter than the WT locus which can be seen from the gel. Electrogram indicates the precise insertion of the hygromycin cassette on the genome. This result indicates that CRISPR/Cas9 system in *L. passim* can be used to replace the entire gene.

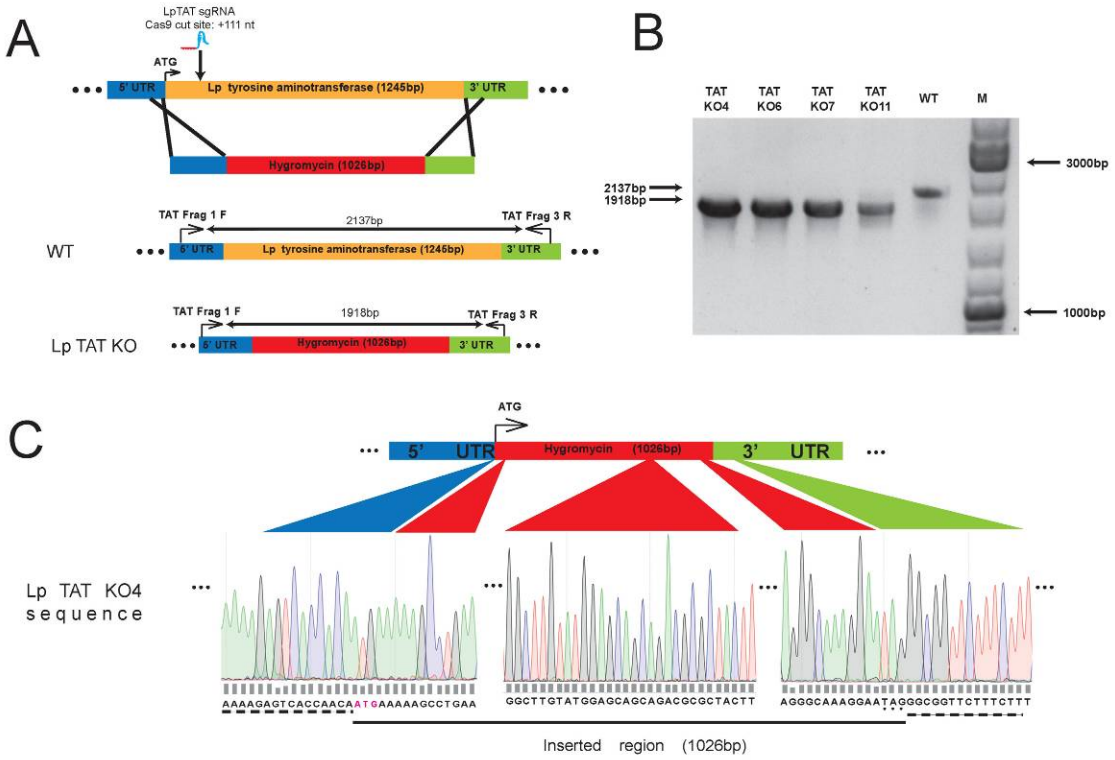


Figure 4.16 Generation of *LpTAT*-KO mutant by CRISPR/Cas9 induced homologous recombination in *L. passim* promastigotes. (A) Schematic representation of the strategy used to generate a *LpTAT*-KO by HDR. The putative cleavage site generated by Cas9 is at +111 of *LpTAT* ORF. The double-stranded break was repaired with a hygromycin B phosphotransferase cassette flanking with ~500-bp homologous regions from the *LpTAT* cleavage site. Schematic representation of the molecular constructs of *LpTAT* WT and KO strains. Primers (arrows) that were used to verify gene disruption by PCR. (B) PCR products obtained with genomic DNA from different *LpTAT* KO parasite lines using upstream and downstream primer sets were analyzed by electrophoresis with 0.8% agarose gel. The *LpTAT* KO strains generated a band with the size of 1,918 bp whereas the *L. passim* WT had a band with the size of 2,137 bp. Lanes: TAT KO4, *LpMT* KO strain 4; TAT KO6, *LpTAT* KO strain 6; TAT KO7, *LpTAT* KO strain 7; TAT KO11, *LpTAT* KO strain 11; WT, *L. passim* wild type; M: 100 bp to 10 kb molecular ladder. (C) Electrogram showing precise insertion of the hygromycin B phosphotransferase cassette into the Cas9 cleavage site by HDR of *LpTAT* KO strain 4. The start codon and the stop codon of hygromycin B phosphotransferase ORF are shown respectively.

4.3.5.9. *L. passim* miltefosine transporter (*LpMT*) KO by CRISPR/Cas9 with hygromycin-containing cassette

I also used *LpMT* as a target since this gene was successfully knocked out in *L. donovani* by CRISPR/Cas9 system (Zhang and Matlashewski, 2015). Figure 4.17 demonstrates that the downstream homologous region was located immediately after the putative cleavage site. For each KO strain, it is difficult to distinguish by the length of the amplified region. Nevertheless, after sequencing the PCR products from these strains, the hygromycin cassette was correctly inserted into the cleavage site of *LpMT* gene. All of these results indicate that gene disruption can be achieved in *L. passim* by CRISPR/Cas9 system.

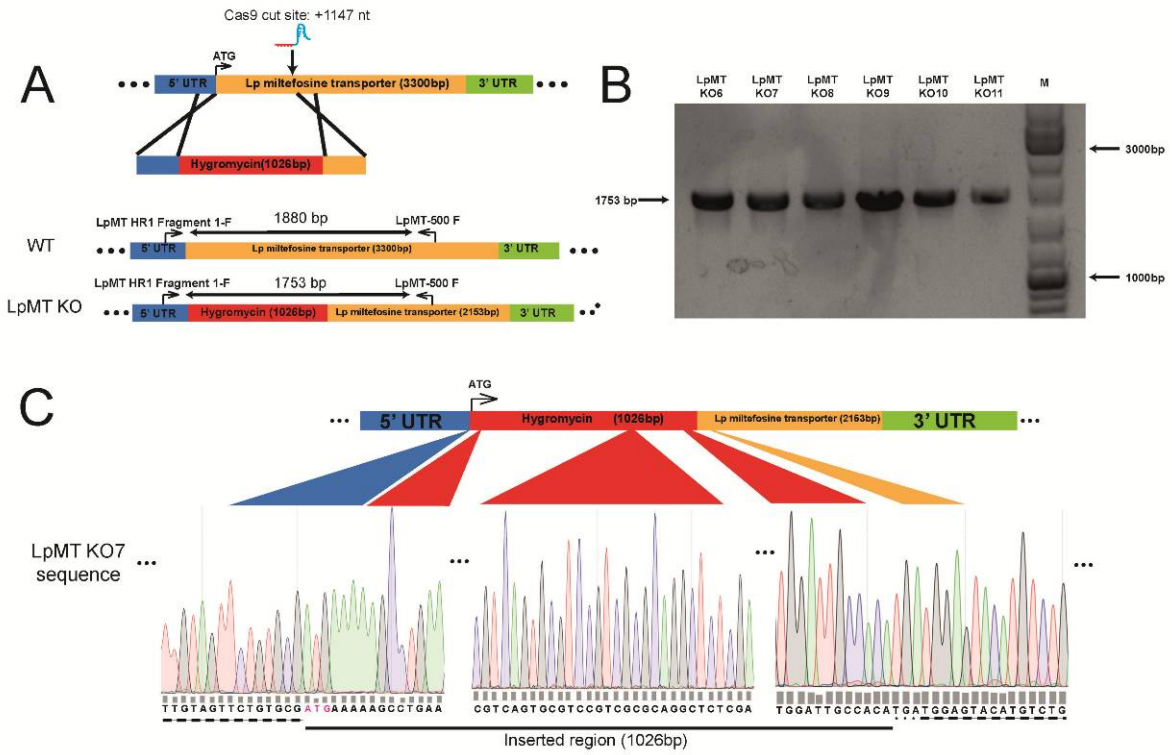


Figure 4.17 Generation of *LpMT*-KO mutant by CRISPR/Cas9 induced homologous recombination in *L. passim* promastigotes. (A) Schematic representation of the strategy used to generate a *LpMT*-KO by HDR. The putative cleavage site generated by Cas9 is at +1,147 of *LpMT* ORF. The double-stranded break was repaired with a hygromycin B phosphotransferase cassette flanking ~500-bp homologous regions from the *LpMT* cleavage site. Schematic representation of the molecular constructs of *LpMT* WT and KO strains. Primers (arrows) that were used to verify gene disruption by PCR. (B) PCR products obtained with genomic DNA from different *LpMT* KO parasite lines using upstream and downstream primer sets were analyzed by electrophoresis with 0.8% agarose gel. Lanes: *LpMT* KO 6-11, *LpMT* KO strain 6-11; M: 100 bp to 10 kb molecular ladder. (C) Electrogram showing precise insertion of the hygromycin B phosphotransferase cassette into the Cas9 cleavage site by HDR of *LpMT* KO strain 7. The start codon and the stop codon of hygromycin B phosphotransferase ORF are shown respectively.

4.3.6. *LpMT* and *LpTAT* homozygous KOs by CRISPR/Cas9 do not express the mRNAs

After co-transfection, the transfectants were maintained in the medium containing three antibiotics. The heterozygous enriched population can be obtained first. At this stage, the selective marker was integrated into one allele. Since *L. passim* is likely to be diploid as predicted by Companion in the previous chapter, the other allele was still intact WT. Later on, in order to facilitate the activity of CRISPR/Cas9 system, antibiotics with higher concentration were added. Then the cell population starts transit from heterozygous to homozygous KO cells which can be detected by PCR since no WT allele can be found anymore. I tested *LpMT* and *LpTAT* mRNA expression in the homozygous-, heterozygous- KO strains together with WT. Figure 4.18 shows there is no mRNA expression of target gene in the homozygous KO strain. The heterozygous KO and WT have the expression of target genes.

I also analyzed the growth rates of these two homozygous KO strains in the culture medium. *LpMT* and *LpTAT* homozygous KO grow slower than *L. passim* WT. This indicates that *LpMT* and *LpTAT* may play another role such as nutrient acquisition and the metabolism which can affect the growth.

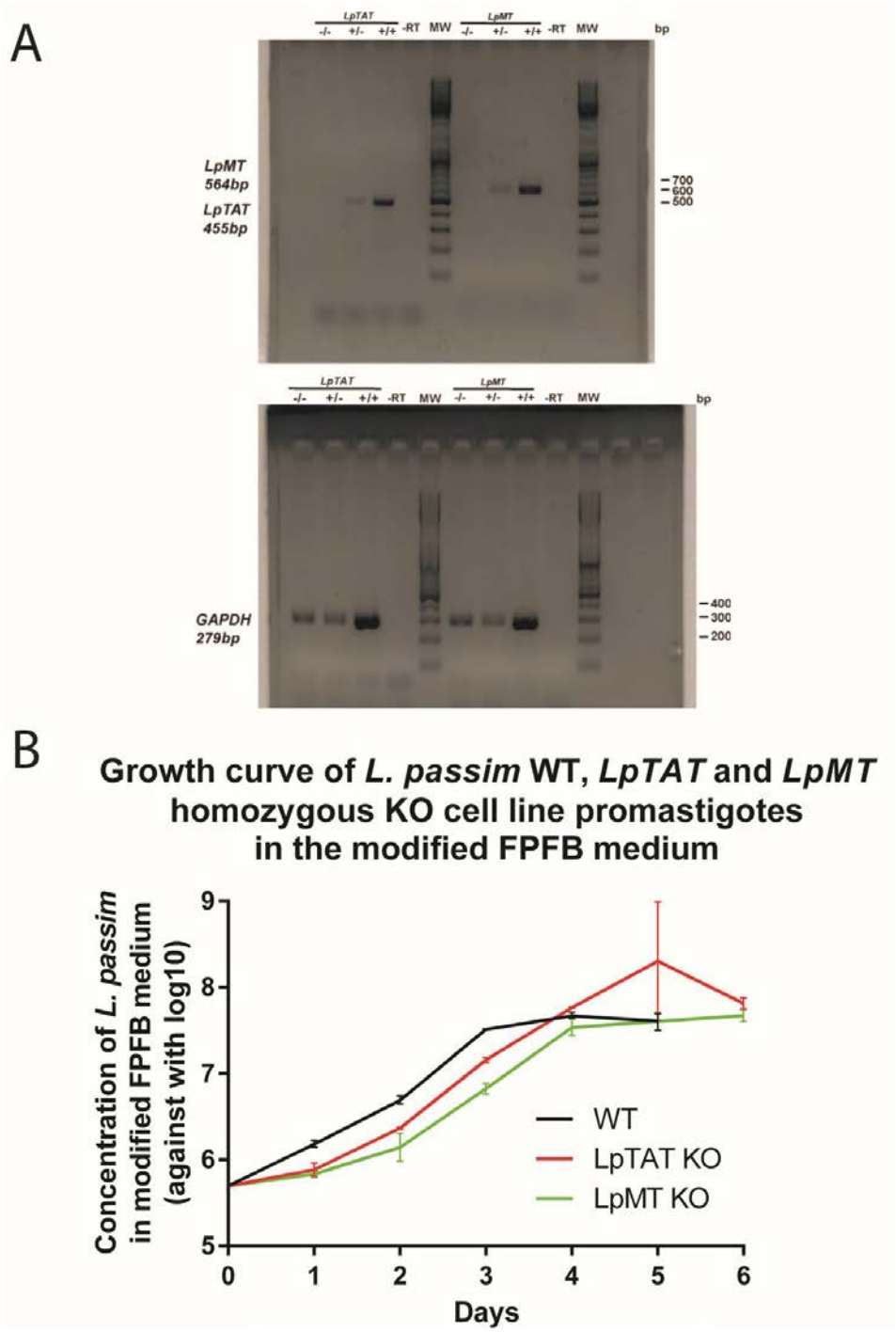


Figure 4.18 Characterization of *LpTAT* and *LpMT* KO strains. (A) Expression of *LpTAT* and *LpMT* mRNAs in the heterozygous- or homozygous-KO strains together with wild type tested by RT-PCR. *LpTAT* heterozygous KO and WT have bands with the expected size of 455 bp. *LpMT* heterozygous and WT have bands with the expected size of 564 bp. These bands are absent in *LpMT* and *LpTAT* homozygous KO strains. The *L. passim* GAPDH gene was used as a positive control which as an expected size of 279 bp. The negative control was run using water as the template for PCR. (B) Growth rates of *LpTAT* and *LpMT* homozygous KO strains together with *L. passim* WT in the modified FPFB medium. Different colors

indicate different *L. passim* parasite lines. Two different clones were tested for each homozygous KO strain. The *L. passim* WT strain was performed in triplicate. The error bars indicate the standard deviation at each time point. *LpTAT* -/- KO: *L. passim* tyrosine aminotransferase homozygous knock out; *LpTAT* +/- KO: *L. passim* tyrosine aminotransferase heterozygous knock out; *LpMT* -/- KO: *L. passim* miltefosine transporter homozygous knock out; *LpMT* +/- KO: *L. passim* miltefosine transporter heterozygous knock out; WT: *L. passim* wild type; -RT: negative control; M: molecular marker.

4.3.7. Comparison of gene knock-out by homologous recombination and CRISPR/Cas9

The results of the previous section demonstrate that there is no change around the putative cleavage site in *LpMT* genomic sequence by expressing only Cas9 and sgRNA. Thus, I wanted to address the question if the homozygous KO was achieved by CRISPR/Cas9 system or the classical HR. After the transfection, the transfectants were maintained under the same concentration of hygromycin and the medium was replaced simultaneously. In the CRISPR/Cas9 KO strain, G418 and blasticidin were added in order to facilitate the expression of sgRNA and Cas9, respectively. After 2 months' selection, the individual clones were isolated and expanded in 96-well plate by limiting dilution. 10 individual expanded clones were randomly selected for DNA extraction and the genomic DNA was used for PCR. In figure 4.19, HDR induced by CRISPR/Cas9 correctly inserted the hygromycin cassette in *LpMT* gene near the expected cleavage site. Furthermore, no WT allele can be detected, indicating that these colonies were homozygous KO. With the classical HR method, I found all clones have the wild type allele by genomic PCR. However, genomic PCR to detect the KO allele by homologous recombination reveals that 7/10 colonies have the expected insertion of the hygromycin cassette. For the three clones, hygromycin cassette should be inserted to other region of the *L. passim* genome. Above results strongly indicate it is possible to obtain homozygous KO using single antibiotic-resistance cassette by CRISPR/Cas9 system but not the classical HR method.

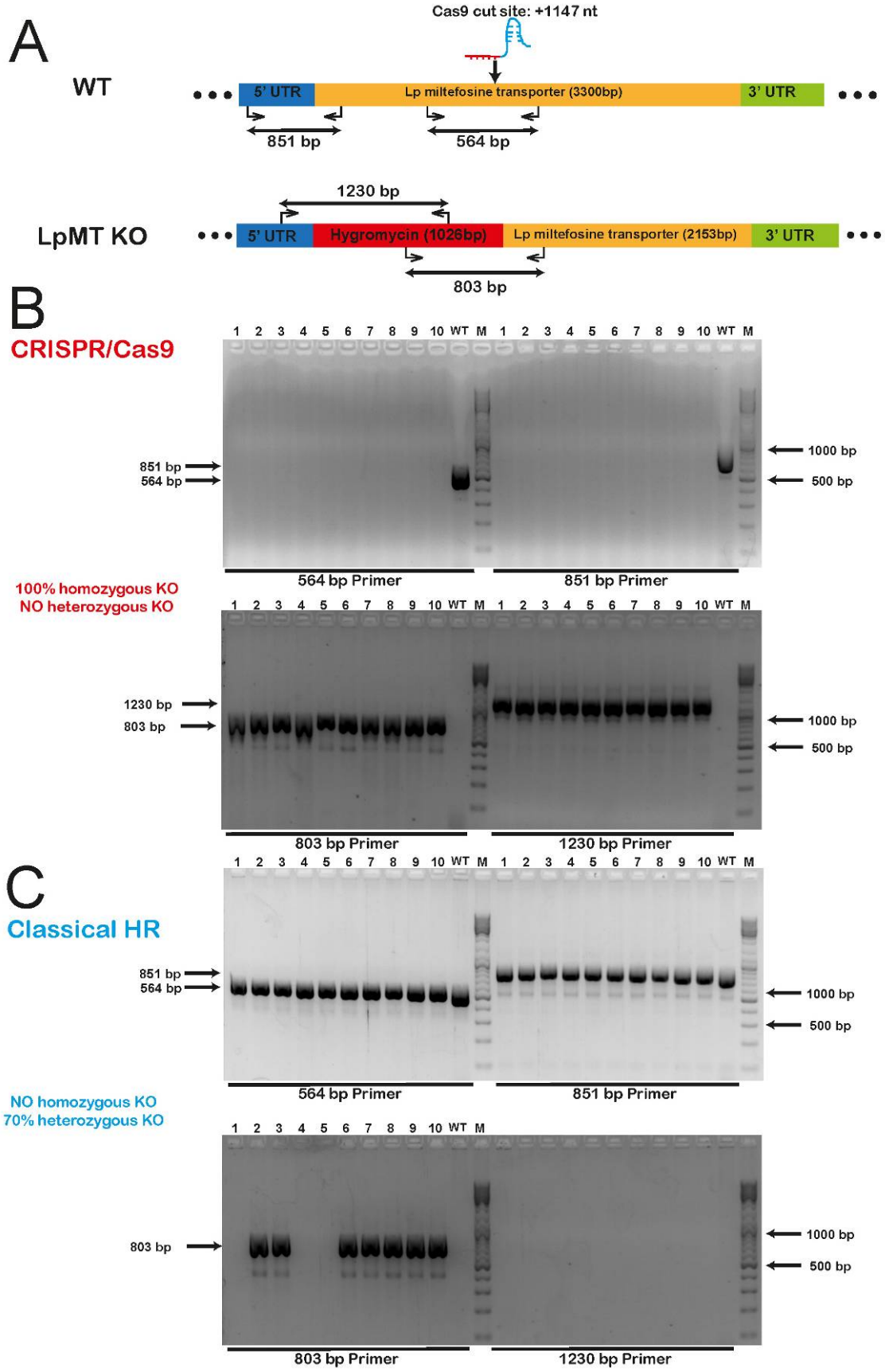


Figure 4.19 Comparison of gene KO by CRISPR/Cas9-induced HDR method and the classical homologous recombination (HR) method. (A) Schematic representation of WT allele and *LpMT*-KO allele Primers (arrows) and the expected amplicon size are shown to verify the WT allele and *LpMT*-KO allele. (B) PCR products obtained with genomic DNA from 10

randomly picked single colonies using CRISPR/Cas9 method. (C) PCR products obtained with genomic DNA from 10 randomly picked single colonies using the classical HR method. The homozygous and heterozygous KO rates are shown on the left. The primer sets used are shown at the bottom of each gel and the expected sizes of amplicons are shown on the left whereas the positions of 500 and 1000 bp bands are shown on the right. WT: *L. passim* WT; M: 10 kb molecular ladder.

4.4. Discussion

Understanding the host-parasite interactions is significantly facilitated by the availability of genome sequences of both honey bee and *L. passim*. Also, the application of molecular tools makes it possible to understand the gene functions. As far as I know, my work is the first one to establish CRISPR/Cas9 system in this parasite, *L. passim*. Previous reports on *T. cruzi* and *Chlamydomonas reinhardtii* indicate that the expression of Cas9 is toxic since it suppressed the growth of cells and the cells tend to lose the expression of Cas9 (Jiang et al., 2014;Peng et al., 2015). However, Cas9 was well-tolerated by *L. passim*. Similarly, Cas9 was also well-tolerated by *L. mexicana* as reported previously (Beneke et al., 2017). Due to *L. passim* could lose the expression of exogenous plasmids as I showed above, for the stable Cas9-expressing parasite line, it is recommended to add antibiotics (in this case, blasticidin) to maintain the expression of Cas9 protein. Furthermore, this also indicates that *T. cruzi* ribosomal promoter can be recognized and utilized by *L. passim*. From the growth curve of both WT and Cas9 expressing parasite line, it takes 2 days until both strains reach to the mid-log phase and 4 days until both strains reach to the stationary phase. For the CRISPR-1V strategy (one vector for expressing Cas9 and sgRNA), sgRNA/Cas9/pTREX-n plasmid was used for transfection due to the fact that transcription in trypanosomatids is polycistronic. However, after several days of selection with antibiotics, there is no green fluorescence can be observed under the microscope when using the CRISPR-1V strategy. Thus, by using CRISPR-2V strategy, I achieved the goal by using CRISPR/Cas9 system to knock out the endogenous genes. Two genes were replaced by the drug resistance marker as in *T. cruzi* and *L. donovani* (Lander et al., 2015a;Zhang and Matlashewski, 2015). I obtained heterozygous first and these heterozygous KO cells tend to transform into homozygous by CRISPR/Cas9. This indicates in CRISPR/Cas9 in *L. passim* tends to cleave one allele and repair the DSB with the presence of the homologous template. At this stage, the transfectants are heterozygous. After selection with a higher concentration of antibiotics, the CRISPR/Cas9 system might targets to another allele which leads to the homozygous KO. Furthermore, since mRNAs of these target genes were absent, the expression of these target proteins must be eliminated. All these results

indicate CRISPR/Cas9 can be used to knock out genes in *L. passim* and help us to identify the gene functions. Meanwhile, I also compared the traditional homologous recombination method with CRISPR/Cas9 method. This traditional method usually requires two rounds of transfection with two antibiotics resistance genes in order to replace two alleles. This procedure usually has low efficiency and is time-consuming (Santrich et al., 1997). From our results, after two months' selection with antibiotics, the colonies I obtained were all homozygous KO. As a comparison, using classical HR method I obtained no homozygous KO and 70% of the selected clones were the heterozygous KO. At last, it should be noticed that in all clones using classical HR method, the 5' UTR of the hygromycin gene is missing. Although the forward primer of this 1,230 bp region is outside the homologous arm, it is only 19 bp upstream of the homologous region. This may induce some small changes near this region which may result in the failure of the primer binding. Notably, my result here may also indicate a drawback of genome editing when using classical HR method since it may cause unexpected genome changes. I observed that in the absence of a donor DNA, when targeting to endogenous genes, *L. passim* genome remains intact with no MMEJ and chromosomal translocation detected. Also, I used T7EI to detect if there are small indels near the cleavage sites. Unexpectedly, all these results indicate using sgRNA expression vector only cannot trigger either MMEJ or chromosomal translocation in *L. passim*. I intend to express two sgRNAs targeting to *LpMT* gene in the future to investigate more about the mechanism of CRISPR/Cas9 in *L. passim*. For endogenous genes, I may assume that with sgRNA expression only, *L. passim* repairs perfectly the DSB induced by CRISPR/Cas9 system. This phenomenon has never been found in any other species. Since this species is recently classified and our phylogenetic analysis also indicates that *Lotmaria* is distinct from Trypanosome, Leishmania, Leptomonas and Crithidia. This may indicate Lotmaria spp. may have evolved a different pathway in order to perfectly repair the DSB by CRISPR/Cas9 system. Also, for the endogenous *LpMT* gene, there are micro-homology sequences (GTGGA) near the putative cleavage site in the genome. However, I did not observe any indels near the cleavage site. The mechanism of how the parasites use the specific micro-homology sequences is still unknown. However, when the homologous cassette is present together with sgRNA expression vector, our results indicate a clear insertion of the donor DNA in the desired gene targets by CRISPR/Casa9 induced homologous recombination. I obtained the heterozygous KO in 15 days after selection in the medium. In order to further obtain the

homozygous KO, I maintained the antibiotics in the medium for another 4 weeks. Earlier time points can be tested in order to identify homozygous KO cells. From my observation, a mixture of hete- and homozygous KO were obtained first. Later on, only homozygous KO can be found. However, the specific mechanism of this transition is not well understood yet. One possible explanation for this might be the Cas9 may randomly cleavage one allele or two alleles which result in the generation of hete- and homozygous KO cells. Later, although the heterozygous cells are hygromycin resistant, they might be eliminated by a high concentration of hygromycin compared to the homozygous KO cells. The other mechanism is somehow similar to mutagenic chain reaction (MCR) as reported recently which can be used to convert heterozygous to homozygous KO (Gantz and Bier, 2015). Similarly, the same phenomenon was also observed when CRISPR/Cas9 was applied in *T. cruzi* (Lander et al., 2015a). From the *L. passim* annotation and protein coding sequences, there are several RAD DNA repair proteins can be found. Such as: DNA repair protein RAD51 (Lp_280006800.1), DNA repair protein RAD2 (Lp_350029200.1), DNA repair protein Rad9 (Lp_150009500.1), DNA repair and recombination protein RAD54 (Lp_240008400.1), double-strand-break repair protein rad21 homolog (Lp_050017800.1), RAD50 DNA repair-like protein (Lp_280006300.1) and recombinase rad51 (Lp_000220500.1, Lp_330022700.1). Also, another important factor meiotic recombination protein DMC1 (Lp_350038300.1) can also be found in the *L. passim* protein-coding and annotation files. Moreover, three BRCA1-related genes also can be found in the *L. passim* genome, which are: Pescadillo N-terminus/BRCA1 C Terminus (BRCT) domain containing protein (Lp_040007000.1), DNA repair protein BRCA2 (Lp_200005700.1) and BRCA1 C Terminus (BRCT) domain containing protein (Lp_350037500.1). *T. brucei* and *T. cruzi* Dmc1 genes have 65% and 70% identity to human DMC1 genes (Proudfoot and McCulloch, 2006;Kelso et al., 2017a). There is only one DMC1 gene among the *L. passim* protein coding sequences (Lp_350038300.1 meiotic recombination protein DMC1) which has 79% of the identity compared with the human DMC1 RefSeqGene on chromosome 22 (NG_017203.1) using tBLSATn. All these genes above might involve in the DSB repair pathway by HDR. *L. passim* also contains several critical components for the MMEJ pathway. As reviewed before, these genes include flap endonuclease-1 (FEN-1) (Lp_000347800.1) (Sharma et al., 2016), poly (ADP-ribose) polymerase 1 (PARP1), DNA polymerase θ and Mre11-Rad50-Xrs2 (MRX) complex (Seol et al., 2017). However, take *L. passim* PARP1 gene (Lp_080009900.1) as an example: it only shares 48% of the identity with the PARP1 gene in *T.*

brucei by using tBLASTn algorithm. As for DNA polymerase θ, *L. passim* only codes the helicase domain (Lp_000041700.1) and the polymerase domain (Lp_000227200.1). Since the DNA polymerase θ protein contains three major domains which are helicase-like domain, central domain and the polymerase domain (Sfeir and Symington, 2015;Wang and Xu, 2017). Specifically, as for the polymerase domain, it contains three highly conserved loops which are critical for MMEJ especially for the loop 2 (55 amino acids) (Kent et al., 2015). However, this loop 2 region seems missing in the *L. passim* polymerase domain (Lp_000227200.1) when using BLASTp algorithm which may indicate the dysfunction of this protein domain due to the missing of essential region. In addition, only two of them can be found in the protein coding sequences. The central domain cannot be found in the predicted protein. The functions of *L. passim* DNA polymerase θ could be different from those of other species because the central domain is missing. In *Saccharomyces cerevisiae*, the MRX complex was shown to be important in MMEJ which initiate the end resection, the first step in MMEJ. The endo/exonuclease Mre11 (Lp_270017800.1) and Rad50 DNA repair-like protein (Lp_280006300.1) are present in *L. passim* genome. However, *L. passim* does not code the Xrs2. This might be the reason why MMEJ pathway was absent since the MRX complex may not function without the Xrs2 component. Another critical gene for MMEJ pathway in the yeast, Sae2, is also missing in *L. passim* genome (Sfeir and Symington, 2015). Thus, due to the low degree of conservation of some essential genes and some core components missing from *L. passim* genome, MMEJ pathway could be absent in *L. passim* as observed in my experiments. Furthermore, *L. passim* RAD51 inhibitors such as B02 and RI-1 might be applied to inhibit the HDR pathway in order to stimulate the potential MMEJ pathway as applied in *L. donovani* (Zhang et al., 2017).

Since using sgRNA expression vector only did not show any changes in the genome when targeting endogenous genes of *L. passim*, the best strategy to establish CRISPR/Cas9 system in *L. passim* would be to introduce the sgRNA expression plasmid and donor DNA containing drug resistance markers to the Cas9 stable expression parasites. However, the CRISPR/Cas9 system is not invincible when manipulating the genome. For example, the CRISPR/Cas9 system cannot be used to KO genes if they are essential for cells to survive. Under this circumstance, the second allele can never be replaced after a long-time selection. This phenomenon has been found in *L. donovani* from the previous report (Zhang and Matlashewski, 2015). In *L. passim* CRISPR/Cas9 system I established here, some

improvements can be done in the future. Firstly, since promoters to express sgRNA and Cas9 are from either Leishmania or Trypanosome, it is better to use *L. passim* specific promoter sequences to maximize the expression. Secondly, the recent study in *L. mexicana* has indicated using CRISPR/Cas9 system can significantly decrease the required length of homologous arms from 350 nt to only 24 nt (Beneke et al., 2017). I may assume that short homologous flanks can also be correctly integrated by using *L. passim* CRISPR/Cas9 system. Under this circumstance, the homologous flanks containing drug resistance gene can be amplified in one step with high accuracy since only the plasmid is needed. Thirdly, N/C-terminal tagging may also be performed in the future by using CRISPR/Cas9 in order to introduce a fluorescence marker. The fluorescent marker can be used by FACS sorting in order to select the transfected parasites. At last, I would encourage researchers to improve the genome assembly of *L. passim*. From our RNA-seq results, it would be promising to KO two gp63 genes since these two genes were upregulated by infection in honey bee gut. However, due to the insufficient information on the contigs containing GP63 genes, I was unable to amplify the homologous arms. Improving the genome assembly will provide more information of genome and it may also become possible to predict the potential chromosomal translocation (if any) on the genome by introducing the oligonucleotide as shown in *L. donovani* (Zhang et al., 2017).

Although the specific DSB repair mechanism in *L. passim* needs to be further investigated, our CRISPR/Cas9 system can be used to facilitate the understanding of gene functions in this newly classified parasite. From our RNA-seq result, several genes might be good targets to KO which may play critical roles during infection. For example, *L. passim* Asparaginase (Lp_000354500.1) gene was significantly upregulated at all time points during infection. This gene was proposed to have a neutralizing function which can adjust the pH level in the gut for *Leishmania* in order to survival in the host (Singh et al., 2015). Another promising target is *L. passim* paraflagellar rod protein 5 (Lp_270017400.1) which downregulated at all time points during the infection. The downregulation of this gene may due to the transformation of *L. passim* from the swimming to attached form since *L. passim* do not need the flagellum after they attach to the honey bee hindgut. KO of this gene may result in the immobility of *L. passim* promastigotes as previously shown in *T. cruzi* (Lander et al., 2015b). Several upregulated genes that related to peptidase activity are also good targets for KO. For example, *L. passim* carboxypeptidase (Lp_000297100.1) and two copies of *L. passim* GP63,

leishmanolysin (Lp_100006400.1, Lp_000433700.1) are good candidates for KO. After knocking out these promising genes, these strains can be fed to the honey bee in order to test if the effects of gene KO for the infection. Meanwhile, these genes and their orthologs may also be good candidates in medically important parasites such as *Leishmania* and *Trypanosomes*. All in all, this gene manipulation method may help us to control the parasite on molecular levels which could protect foragers in the future.

5. Final conclusion

In this project, the effects of *L. passim* infection under both laboratory condition and natural condition were analyzed in deeper ways. Also, with the help of RNA-seq, genes differentially expressed from both *L. passim* and honey bee were analyzed at different time points. This helps us to understand the host-parasite interaction better on a molecular level. Besides, this is the first time that CRISPR/Cas9 system was successfully established in *L. passim*. This will facilitate the understanding of the genes of this parasite in the future.

6. References

Aizen, M.A., Garibaldi, L.A., Cunningham, S.A., and Klein, A.M. (2008). Long-term global trends in crop yield and production reveal no current pollination shortage but increasing pollinator dependency. *Curr Biol* 18**,** 1572-1575.

Alaux, C., Dantec, C., Parrinello, H., and Le Conte, Y. (2011). Nutrigenomics in honey bees: digital gene expression analysis of pollen's nutritive effects on healthy and varroa-parasitized bees. *BMC genomics* 12**,** 496.

Alberoni, D., Baffoni, L., Gaggìa, F., Ryan, P., Murphy, K., Ross, P., Stanton, C., and Di Gioia, D. (2018). Impact of beneficial bacteria supplementation on the gut microbiota, colony development and productivity of Apis mellifera L. *Beneficial microbes* 9**,** 269-278.

Alonso, V.L., Ritagliati, C., Cribb, P., and Serra, E.C. (2014). Construction of three new Gateway(R) expression plasmids for Trypanosoma cruzi. *Mem Inst Oswaldo Cruz* 109**,** 1081-1085.

Amdam, G.V., Hartfelder, K., Norberg, K., Hagen, A., and Omholt, S.W. (2004). Altered physiology in worker honey bees (Hymenoptera: Apidae) infested with the mite Varroa destructor (Acari: Varroidae): a factor in colony loss during overwintering? *Journal of economic entomology* 97**,** 741-747.

Amdam, G.V., and Omholt, S.W. (2003). The hive bee to forager transition in honeybee colonies: the double repressor hypothesis. *Journal of theoretical biology* 223**,** 451-464.

Amdam, G.V., Simões, Z.L., Guidugli, K.R., Norberg, K., and Omholt, S.W. (2003). Disruption of vitellogenin gene function in adult honeybees by intra-abdominal injection of double-stranded RNA. *BMC biotechnology* 3**,** 1.

Ament, S.A., Chan, Q.W., Wheeler, M.M., Nixon, S.E., Johnson, S.P., Rodriguez-Zas, S.L., Foster, L.J., and Robinson, G.E. (2011). Mechanisms of stable lipid loss in a social insect. *Journal of Experimental Biology* 214**,** 3808-3821.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31**,** 166-169.

Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome research* 22**,** 2008-2017.

Anderson, K., Sheehan, T., Eckholm, B., Mott, B., and Degrandi-Hoffman, G. (2011). An emerging paradigm of colony health: microbial balance of the honey bee and hive (Apis mellifera). *Insectes Sociaux* 58**,** 431-444.

Antúnez, K., Martín-Hernández, R., Prieto, L., Meana, A., Zunino, P., and Higes, M. (2009). Immune suppression in the honey bee (Apis mellifera) following infection by Nosema ceranae (Microsporidia). *Environmental microbiology* 11**,** 2284-2290.

Arismendi, N., Bruna, A., Zapata, N., and Vargas, M. (2016). PCR-specific detection of recently described Lotmaria passim (Trypanosomatidae) in Chilean apiaries. *J Invertebr Pathol* 134**,** 1-5.

Aufauvre, J., Misme-Aucouturier, B., Viguès, B., Texier, C., Delbac, F., and Blot, N. (2014). Transcriptome analyses of the honeybee response to Nosema ceranae and insecticides. *PLoS One* 9**,** e91686.

Aylon, Y., Liefshitz, B., and Kupiec, M. (2004). The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle. *The EMBO journal* 23**,** 4868-4875.

Barnes, R.L., and Mcculloch, R. (2007). Trypanosoma brucei homologous recombination is dependent on substrate length and homology, though displays a differential dependence on mismatch repair as substrate length decreases. *Nucleic acids research* 35**,** 3478-3493.

Beneke, T., Madden, R., Makin, L., Valli, J., Sunter, J., and Gluenz, E. (2017). A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. *Open Science* 4**,** 170095.

Benítez, D., Medeiros, A., Fiestas, L., Panozzo-Zenere, E.A., Maiwald, F., Prousis, K.C., Roussaki, M., Calogeropoulou, T., Detsi, A., and Jaeger, T. (2016). Identification of novel chemical scaffolds inhibiting trypanothione synthetase from pathogenic trypanosomatids. *PLoS neglected tropical diseases* 10**,** e0004617.

Benoit, J.B., Vigneron, A., Broderick, N.A., Wu, Y., Sun, J.S., Carlson, J.R., Aksoy, S., and Weiss, B.L. (2017). Symbiont-induced odorant binding proteins mediate insect host hematopoiesis. *Elife* 6**,** e19535.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., and Haas, B. (2005). The genome of the African trypanosome Trypanosoma brucei. *science* 309**,** 416-422.

Bitondi, M., and Simoes, Z.P. (1996). The relationship between level of pollen in the diet, vitellogenin and juvenile hormone titres in Africanized Apis mellifera workers. *Journal of Apicultural Research* 35**,** 27-36.

Blackmore, S., Wortley, A.H., Skvarla, J.J., and Rowley, J.R. (2007). Pollen wall development in flowering plants. *New Phytologist* 174**,** 483-498.

Bottacini, F., Milani, C., Turroni, F., Sánchez, B., Foroni, E., Duranti, S., Serafini, F., Viappiani, A., Strati, F., and Ferrarini, A. (2012). Bifidobacterium asteroides PRL2011 genome analysis reveals clues for colonization of the insect gut. *PLoS One* 7**,** e44229.

Boucher, N., Wu, Y., Dumas, C., Dubé, M., Sereno, D., Breton, M., and Papadopoulou, B. (2002). A Common Mechanism of Stage-regulated Gene Expression inLeishmania Mediated by a Conserved 3′-Untranslated Region Element. *Journal of Biological Chemistry* 277**,** 19511-19520.

Boulanger, N., Bulet, P., and Lowenberger, C. (2006). Antimicrobial peptides in the interactions between insects and flagellate parasites. *Trends in parasitology* 22**,** 262-268.

Boulanger, N., Ehret-Sabatier, L., Brun, R., Zachary, D., Bulet, P., and Imler, J.-L. (2001). Immune response of Drosophila melanogaster to infection with the flagellate parasite Crithidia spp. *Insect biochemistry and molecular biology* 31**,** 129-137.

Boulanger, N., Lowenberger, C., Volf, P., Ursic, R., Sigutova, L., Sabatier, L., Svobodova, M., Beverley, S.M., Späth, G., and Brun, R. (2004). Characterization of a defensin from the sand fly Phlebotomus duboscqi induced by challenge with bacteria or the protozoan parasite Leishmania major. *Infection and Immunity* 72**,** 7140-7146.

Bouvier, J., Schneider, P., Etges, R., and Bordier, C. (1990). Peptide substrate specificity of the membrane-bound metalloprotease of Leishmania. *Biochemistry* 29**,** 10113-10119.

Brown, M.J., Schmid-Hempel, R., and Schmid-Hempel, P. (2003). Strong context-dependent virulence in a host–parasite system: reconciling genetic evidence with theory. *Journal of Animal Ecology* 72**,** 994-1002.

Brutscher, L.M., Daughenbaugh, K.F., and Flenniken, M.L. (2015). Antiviral defense mechanisms in honey bees. *Current opinion in insect science* 10**,** 71-82.

Brutscher, L.M., Daughenbaugh, K.F., and Flenniken, M.L. (2017). Virus and dsRNA-triggered transcriptional responses reveal key components of honey bee antiviral defense. *Sci Rep* 7**,** 6448.

Bucekova, M., Sojka, M., Valachova, I., Martinotti, S., Ranzato, E., Szep, Z., Majtan, V., Klaudiny, J., and Majtan, J. (2017). Bee-derived antibacterial peptide, defensin-1, promotes wound re-epithelialisation in vitro and in vivo. *Scientific reports* 7**,** 7340.

Buttstedt, A., Moritz, R.F., and Erler, S. (2014). Origin and function of the major royal jelly proteins of the honeybee (Apis mellifera) as members of the yellow gene family. *Biological Reviews* 89**,** 255-269.

Cadavid-Restrepo, G., Gastardelo, T.S., Faudry, E., De Almeida, H., Bastos, I.M., Negreiros, R.S., Lima, M.M., Assumpção, T.C., Almeida, K.C., and Ragno, M. (2011). The major leucyl aminopeptidase of Trypanosoma cruzi (LAPTc) assembles into a homohexamer and belongs to the M17 family of metallopeptidases. *BMC biochemistry* 12**,** 46.

Campbell, D.A., Thomas, S., and Sturm, N.R. (2003). Transcription in kinetoplastid protozoa: why be normal? *Microbes and infection* 5**,** 1231-1240.

Canavaci, A.M., Bustamante, J.M., Padilla, A.M., Perez Brandan, C.M., Simpson, L.J., Xu, D., Boehlke, C.L., and Tarleton, R.L. (2010). In vitro and in vivo high-throughput assays for the testing of anti-Trypanosoma cruzi compounds. *PLoS Negl Trop Dis* 4**,** e740.

Cariveau, D.P., Elijah Powell, J., Koch, H., Winfree, R., and Moran, N.A. (2014). Variation in gut microbial communities and its association with pathogen infection in wild bumble bees (Bombus). *Isme j* 8**,** 2369-2379.

Carroll, M.J., Brown, N., Goodall, C., Downs, A.M., Sheenan, T.H., and Anderson, K.E. (2017). Honey bees preferentially consume freshly-stored pollen. *PLoS One* 12**,** e0175933.

Carton, Y., Frey, F., and Nappi, A. (2009). Parasite-induced changes in nitric oxide levels in Drosophila paramelanica. *Journal of Parasitology* 95**,** 1134-1141.

Castelli, L., Branchiccela, B., Invernizzi, C., Tomasco, I., Basualdo, M., Rodriguez, M., Zunino, P., and Antúnez, K. (2018). Detection of Lotmaria passim in Africanized and European honey bees from Uruguay, Argentina and Chile. *Journal of Invertebrate Pathology*.

Castro, H., Romao, S., Carvalho, S., Teixeira, F., Sousa, C., and Tomas, A.M. (2010). Mitochondrial redox metabolism in trypanosomatids is independent of tryparedoxin activity. *PLoS One* 5**,** e12607.

Cavigli, I., Daughenbaugh, K.F., Martin, M., Lerch, M., Banner, K., Garcia, E., Brutscher, L.M., and Flenniken, M.L. (2016). Pathogen prevalence and abundance in honey bee colonies involved in almond pollination. *Apidologie* 47**,** 251-266.

Cepero, A., Ravoet, J., Gómez-Moracho, T., Bernal, J.L., Del Nozal, M.J., Bartolomé, C., Maside, X., Meana, A., González-Porto, A.V., De Graaf, D.C., Martín-Hernández, R., and Higes, M. (2014). Holistic screening of collapsing honey bee colonies in Spain: a case study. *BMC Research Notes* 7**,** 649.

Cersini, A., Antognetti, V., Conti, R., Velletrani, F., and Formato, G. (2015). First PCR isolation of Crithidia mellificae (Euglenozoa: Trypanosomatidae) in Apis mellifera (Hymenoptera: Apidae) in Italy. *Fragmenta entomologica* 47**,** 45-49.

Chakrabarti, S., Liehl, P., Buchon, N., and Lemaitre, B. (2012). Infection-induced host translational blockage inhibits immune responses and epithelial renewal in the Drosophila gut. *Cell host & microbe* 12**,** 60-70.

Chang, C.-I., Chelliah, Y., Borek, D., Mengin-Lecreulx, D., and Deisenhofer, J. (2006). Structure of tracheal cytotoxin in complex with a heterodimeric pattern-recognition receptor. *Science* 311**,** 1761-1764.

Chiurillo, M.A., Lander, N., Bertolini, M.S., Storey, M., Vercesi, A.E., and Docampo, R. (2017). Different roles of mitochondrial calcium uniporter complex subunits in growth and infectivity of Trypanosoma cruzi. *MBio* 8**,** e00574-00517.

Clayton, C.E. (2002). Life without transcriptional control? From fly to man and back again. *The EMBO journal* 21**,** 1881-1888.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21**,** 3674-3676.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., and Marraffini, L.A. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339**,** 819-823.

Corona, M., Velarde, R.A., Remolina, S., Moran-Lauter, A., Wang, Y., Hughes, K.A., and Robinson, G.E. (2007). Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. *Proceedings of the National Academy of Sciences* 104**,** 7128-7133.

Corrêa-Da-Silva, M.S., Fampa, P., Lessa, L.P., Dos Reis Silva, E., Dos Santos Mallet, J.R., Saraiva, E.M., and Motta, M.C.M. (2006). Colonization of Aedes aegypti midgut by the endosymbiont-bearing trypanosomatid Blastocrithidia culicis. *Parasitology research* 99**,** 384-391.

Crailsheim, K. (1988). Intestinal transport of sugars in the honeybee (Apis mellifera L.). *Journal of insect physiology* 34**,** 839-845.

Crailsheim, K. (1990). The protein balance of the honey bee worker. *Apidologie* 21**,** 417-429.
Cremonz, T.M., De Jong, D., and Bitondi, M.M. (1998). Quantification of hemolymph proteins as a fast method for testing protein diets for honey bees (Hymenoptera: Apidae). *Journal of economic entomology* 91**,** 1284-1289.
D'avila-Levy, C., Araújo, F., Vermelho, A., Soares, R., Santos, A., and Branquinha, M. (2005). Proteolytic expression in Blastocrithidia culicis: influence of the endosymbiont and similarities with virulence factors of pathogenic trypanosomatids. *Parasitology* 130**,** 413-420.
D'avila-Levy, C.M., Souza, R.F., Gomes, R.C., Vermelho, A.B., and Branquinha, M.H. (2003). A metalloproteinase extracellularly released by Crithidia deanei. *Canadian journal of microbiology* 49**,** 625-632.
D'avila-Levy, C.M., Melo, A.C., Vermelho, A.B., and Branquinha, M.H. (2001). Differential expression of proteolytic enzymes in endosymbiont-harboring Crithidia species. *FEMS microbiology letters* 202**,** 73-77.
Da Silva, F.M., Noyes, H., Campaner, M., Junqueira, A., Coura, J., Añez, N., Shaw, J., Stevens, J., and Teixeira, M. (2004). Phylogeny, taxonomy and grouping of Trypanosoma rangeli isolates from man, triatomines and sylvatic mammals from widespread geographical origin based on SSU and ITS ribosomal sequences. *Parasitology* 129**,** 549-561.
Daniels, J.-P., Gull, K., and Wickstead, B. (2010). Cell biology of the trypanosome genome. *Microbiology and Molecular Biology Reviews* 74**,** 552-569.
Das, S., and Dimopoulos, G. (2008). Molecular analysis of photic inhibition of blood-feeding in Anopheles gambiae. *BMC physiology* 8**,** 23.
Davies, S.-A., and Dow, J.A. (2009). Modulation of epithelial innate immunity by autocrine production of nitric oxide. *General and comparative endocrinology* 162**,** 113-121.
De Freitas Nascimento, J., Kelly, S., Sunter, J., and Carrington, M. (2018). Codon choice directs constitutive mRNA levels in trypanosomes. *Elife* 7**,** e32467.
De Gregoris, T.B., Aldred, N., Clare, A.S., and Burgess, J.G. (2011). Improvement of phylum- and class-specific primers for real-time PCR quantification of bacterial taxa. *Journal of microbiological methods* 86**,** 351-356.
De Melo, A.C.N., D'avila-Levy, C.M., Dias, F.A., Armada, J.L.A., Silva, H.D., Lopes, A.H., Santos, A.L., Branquinha, M.H., and Vermelho, A.B. (2006). Peptidases and gp63-like proteins in Herpetomonas megaseliae: possible involvement in the adhesion to the invertebrate host. *International journal for parasitology* 36**,** 415-422.
De Melo, A.C.N., D'avila-Levy, C.M., Branquinha, M.H., and Vermelho, A.B. (2002). Crithidia guilhermei: gelatin-and haemoglobin-degrading extracellular metalloproteinases. *Experimental parasitology* 102**,** 150-156.
De Melo, A.C.N., Giovanni-De-Simone, S., Branquinha, M.H., and Vermelho, A.B. (2001). Crithidia guilhermei: purification and partial characterization of a 62-kDa extracellular metalloproteinase. *Experimental parasitology* 97**,** 1-8.
Deng, S.K., Gibb, B., De Almeida, M.J., Greene, E.C., and Symington, L.S. (2014). RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. *Nature structural & molecular biology* 21**,** 405.
Deshwal, S., and Mallon, E.B. (2014). Antimicrobial peptides play a functional role in bumblebee anti-trypanosome defense. *Developmental & Comparative Immunology* 42**,** 240-243.

Di Pasquale, G., Salignon, M., Le Conte, Y., Belzunces, L.P., Decourtye, A., Kretzschmar, A., Suchail, S., Brunet, J.-L., and Alaux, C. (2013). Influence of pollen nutrition on honey bee health: do pollen quality and diversity matter? *PloS one* 8**,** e72016.

Dicarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., and Church, G.M. (2013). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res* 41**,** 4336-4343.

Dinglasan, R., Devenport, M., Florens, L., Johnson, J., Mchugh, C., Donnelly-Doman, M., Carucci, D., Yates Iii, J., and Jacobs-Lorena, M. (2009). The Anopheles gambiae adult midgut peritrophic matrix proteome. *Insect biochemistry and molecular biology* 39**,** 125-134.

Docampo, R. (2011). Molecular parasitology in the 21st century. *Essays in biochemistry* 51**,** 1-13.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature biotechnology* 32**,** 1262-1267.

Dos Santos, A.L.S., Abreu, C.M., Batista, L.M., Alviano, C.S., and De Araújo Soares, R.M. (2001). Cell-associated and extracellular proteinases in Blastocrithidia culicis: influence of growth conditions. *Current microbiology* 43**,** 100-106.

Dos Santos, A.L.S., Ferreira, A., Franco, V.A., Alviano, C.S., and De Araújo Soares, R.M. (1999). Characterization of proteinases in Herpetomonas anglusteri and Herpetomonas roitmani. *Current microbiology* 39**,** 61-64.

Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346**,** 1258096.

Durrer, S., and Schmid-Hempel, P. (1994). Shared use of flowers leads to horizontal pathogen transmission. *Proceedings of the Royal Society of London B: Biological Sciences* 258**,** 299-302.

Dwass, M. (1960). "Some k-sample Rank Order Tests, Contributions to Probability and Statistics". Stanford University Press).

El-Sayed, N.M., Myler, P.J., Bartholomeu, D.C., Nilsson, D., Aggarwal, G., Tran, A.-N., Ghedin, E., Worthey, E.A., Delcher, A.L., and Blandin, G. (2005). The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. *Science* 309**,** 409-415.

Ellegaard, K.M., Tamarit, D., Javelind, E., Olofsson, T.C., Andersson, S.G., and Vásquez, A. (2015). Extensive intra-phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut. *BMC genomics* 16**,** 284.

Engel, P., Bartlett, K.D., and Moran, N.A. (2015). The bacterium Frischella perrara causes scab formation in the gut of its honeybee host. *MBio* 6**,** e00193-00115.

Engel, P., Martinson, V.G., and Moran, N.A. (2012). Functional diversity within the simple gut microbiota of the honey bee. *Proceedings of the National Academy of Sciences* 109**,** 11002-11007.

Erler, S., Popp, M., and Lattorff, H.M.G. (2011). Dynamics of immune system gene expression upon bacterial challenge and wounding in a social insect (Bombus terrestris). *PLoS One* 6**,** e18126.

Evans, J., Aronstein, K., Chen, Y.P., Hetru, C., Imler, J.L., Jiang, H., Kanost, M., Thompson, G., Zou, Z., and Hultmark, D. (2006). Immune pathways and defence mechanisms in honey bees Apis mellifera. *Insect molecular biology* 15**,** 645-656.

Evans, J.D., and Schwarz, R.S. (2011). Bees brought to their knees: microbes affecting honey bee health. *Trends in microbiology* 19**,** 614-620.

Fairlamb, A.H., and Cerami, A. (1992). Metabolism and functions of trypanothione in the Kinetoplastida. *Annu Rev Microbiol* 46**,** 695-729.

Fiorillo, A., Colotti, G., Boffi, A., Baiocco, P., and Ilari, A. (2012). The crystal structures of the tryparedoxin-tryparedoxin peroxidase couple unveil the structural determinants of Leishmania detoxification pathway. *PLoS Negl Trop Dis* 6**,** e1781.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *nature* 391**,** 806.

Forêt, S., and Maleszka, R. (2006). Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (Apis mellifera). *Genome research* 16**,** 000-000.

Galbraith, D.A., Yang, X., Niño, E.L., Yi, S., and Grozinger, C. (2015). Parallel epigenomic and transcriptomic responses to viral infection in honey bees (Apis mellifera). *PLoS pathogens* 11**,** e1004713.

Gantz, V.M., and Bier, E. (2015). The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science***,** aaa5945.

Garver, L.S., Xi, Z., and Dimopoulos, G. (2008). Immunoglobulin superfamily members play an important role in the mosquito immune system. *Developmental & Comparative Immunology* 32**,** 519-531.

Gavigan, C.S., Dalton, J.P., and Bell, A. (2001). The role of aminopeptidases in haemoglobin degradation in Plasmodium falciparum-infected erythrocytes. *Molecular and biochemical parasitology* 117**,** 37-48.

Gegear, R.J., Otterstatter, M.C., and Thomson, J.D. (2005). Does parasitic infection impair the ability of bumblebees to learn flower-handling techniques? *Animal Behaviour* 70**,** 209-215.

Gegear, R.J., Otterstatter, M.C., and Thomson, J.D. (2006). Bumble-bee foragers infected by a gut parasite have an impaired ability to utilize floral information. *Proc Biol Sci* 273**,** 1073-1078.

Giacomini, J.J., Leslie, J., Tarpy, D.R., Palmer-Young, E.C., Irwin, R.E., and Adler, L.S. (2018). Medicinal value of sunflower pollen against bee pathogens. *Scientific reports* 8**,** 14394.

Glover, L., Jun, J., and Horn, D. (2010). Microhomology-mediated deletion and gene conversion in African trypanosomes. *Nucleic acids research* 39**,** 1372-1380.

Glover, L., Mcculloch, R., and Horn, D. (2008). Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes. *Nucleic acids research* 36**,** 2608-2618.

Goodwin, S., Mcpherson, J.D., and Mccombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* 17**,** 333.

Goulson, D., Nicholls, E., Botías, C., and Rotheray, E.L. (2015). Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* 347**,** 1255957.

Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., Wildonger, J., and O'connor-Giles, K.M. (2013). Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics* 194**,** 1029-1035.

Hahn, D.A., and Denlinger, D.L. (2011). Energetics of insect diapause. *Annual review of entomology* 56**,** 103-121.

Haile, S., and Papadopoulou, B. (2007). Developmental regulation of gene expression in trypanosomatid parasitic protozoa. *Current opinion in microbiology* 10**,** 569-577.

Haines, L.R., Thomas, J.M., Jackson, A.M., Eyford, B.A., Razavi, M., Watson, C.N., Gowen, B., Hancock, R.E., and Pearson, T.W. (2009). Killing of trypanosomatid parasites by a modified bovine host defense peptide, BMAP-18. *PLoS neglected tropical diseases* 3**,** e373.

Hao, Z., Kasumba, I., and Aksoy, S. (2003). Proventriculus (cardia) plays a crucial role in immunity in tsetse fly (Diptera: Glossinidiae). *Insect biochemistry and molecular biology* 33**,** 1155-1164.

Higes, M., Rodríguez-García, C., Gomez-Moracho, T., Meana Mañes, A., Bartolomé, C., Maside, X., Barrios, L., and Martín Hernández, R. (2016). Survival of honey bees (Apis mellifera) infected with Crithidia mellificae spheroid forms (Langridge and McGhee: ATCC® 30254™) in the presence of Nosema ceranae.

Hillyer, J.F., and Estévez-Lao, T.Y. (2010). Nitric oxide is an essential component of the hemocyte-mediated mosquito immune response against bacteria. *Developmental & Comparative Immunology* 34**,** 141-149.

Hroncova, Z., Havlik, J., Killer, J., Doskocil, I., Tyl, J., Kamler, M., Titera, D., Hakl, J., Mrazek, J., and Bunesova, V. (2015). Variation in honey bee gut microbial diversity affected by ontogenetic stage, age and geographic location. *PloS one* 10**,** e0118707.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* 37**,** 1-13.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4**,** 44.

Huang, S.K., Csaki, T., Doublet, V., Dussaubat, C., Evans, J.D., Gajda, A.M., Gregorc, A., Hamilton, M.C., Kamler, M., and Lecocq, A. (2014). Evaluation of cage designs and feeding regimes for honey bee (Hymenoptera: Apidae) laboratory experiments. *Journal of economic entomology* 107**,** 54-62.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 31**,** 227-229.

Ihle, K.E., Rueppell, O., Huang, Z.Y., Wang, Y., Fondrk, M.K., Page Jr, R.E., and Amdam, G.V. (2015). Genetic architecture of a hormonal response to gene knockdown in honey bees. *Journal of Heredity* 106**,** 155-165.

Ilyasov, R., Gaifullina, L., Saltykova, E., Poskryakov, A., and Nikolaenko, A. (2013). Defensins in the honeybee antiinfectious protection. *Journal of evolutionary biochemistry and physiology* 49**,** 1-9.

Inverso, J.A., Medina-Acosta, E., O'connor, J., Russell, D.G., and Cross, G.A. (1993). Crithidia fasciculata contains a transcribed leishmanial surface proteinase (gp63) gene homologue. *Molecular and biochemical parasitology* 57**,** 47-54.

Iovannisci, D.M., Plested, C.P., and Moe, G.R. (2010). Evidence for rosettes as an unrecognized stage in the life cycle of Leishmania parasites. *J Eukaryot Microbiol* 57**,** 405-414.

Ira, G., Pellicioli, A., Balijja, A., Wang, X., Fiorani, S., Carotenuto, W., Liberi, G., Bressan, D., Wan, L., and Hollingsworth, N.M. (2004). DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1. *Nature* 431**,** 1011.

Isnard, A., Shio, M.T., and Olivier, M. (2012). Impact of Leishmania metalloprotease GP63 on macrophage signaling. *Frontiers in cellular and infection microbiology* 2**,** 72.

Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M.-A., Adlem, E., and Aert, R. (2005). The genome of the kinetoplastid parasite, Leishmania major. *Science* 309**,** 436-442.

Jacobs-Lorena, M., and Oo, M. (1996). The peritrophic matrix of insects. *The biology of disease vectors. University Press of Colorado, Boulder***,** 318-332.

Jaffe, C.L., and Dwyer, D.M. (2003). Extracellular release of the surface metalloprotease, gp63, from Leishmania and insect trypanosomatids. *Parasitology research* 91**,** 229-237.

Jeacock, L., Faria, J., and Horn, D. (2018). Codon usage bias controls mRNA and protein abundance in trypanosomatids. *Elife* 7**,** e32496.

Jeyaprakash, A., Hoy, M.A., and Allsopp, M.H. (2003). Bacterial diversity in worker adults of Apis mellifera capensis and Apis mellifera scutellata (Insecta: Hymenoptera) assessed using 16S rRNA sequences. *J Invertebr Pathol* 84**,** 96-103.

Jiang, W., Brueggeman, A.J., Horken, K.M., Plucinak, T.M., and Weeks, D.P. (2014). Successful transient expression of Cas9 and single guide RNA genes in Chlamydomonas reinhardtii. *Eukaryot Cell* 13**,** 1465-1469.

Johnson, P.J., Kooter, J.M., and Borst, P. (1987). Inactivation of transcription by UV irradiation of T. brucei provides evidence for a multicistronic transcription unit including a VSG gene. *Cell* 51**,** 273-281.

Joshi, P.B., Kelly, B.L., Kamhawi, S., Sacks, D.L., and Mcmaster, W.R. (2002). Targeted gene deletion in Leishmania major identifies leishmanolysin (GP63) as a virulence factor. *Molecular and biochemical parasitology* 120**,** 33-40.

Kamhawi, S. (2006). Phlebotomine sand flies and Leishmania parasites: friends or foes? *Trends Parasitol* 22**,** 439-445.

Kapheim, K.M., Rao, V.D., Yeoman, C.J., Wilson, B.A., White, B.A., Goldenfeld, N., and Robinson, G.E. (2015). Caste-specific differences in hindgut microbial communities of honey bees (Apis mellifera). *PloS one* 10**,** e0123911.

Kelso, A.A., Goodson, S.D., Temesvari, L.A., and Sehorn, M.G. (2017a). Data on Rad51 amino acid sequences from higher and lower eukaryotic model organisms and parasites. *Data in brief* 10**,** 364-368.

Kelso, A.A., Waldvogel, S.M., Luthman, A.J., and Sehorn, M.G. (2017b). Homologous recombination in protozoan parasites and recombinase inhibitors. *Frontiers in microbiology* 8**,** 1716.

Kent, T., Chandramouly, G., Mcdevitt, S.M., Ozdemir, A.Y., and Pomerantz, R.T. (2015). Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase θ. *Nat Struct Mol Biol* 22**,** 230-237.

Kešnerová, L., Mars, R.A., Ellegaard, K.M., Troilo, M., Sauer, U., and Engel, P. (2017). Disentangling metabolic functions of bacteria in the honey bee gut. *PLoS biology* 15**,** e2003467.

Killer, J., Dubná, S., Sedláček, I., and Švec, P. (2014). Lactobacillus apis sp. nov., from the stomach of honeybees (Apis mellifera), having an in vitro inhibitory effect on the causative agents of American and European foulbrood. *International journal of systematic and evolutionary microbiology* 64**,** 152-157.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature methods* 12**,** 357.

Klein, A.-M., Vaissiere, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., and Tscharntke, T. (2007). Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society of London B: Biological Sciences* 274**,** 303-313.

Koch, H., and Schmid-Hempel, P. (2011). Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *Proc Natl Acad Sci U S A* 108**,** 19288-19292.

Koch, H., and Schmid-Hempel, P. (2012). Gut microbiota instead of host genotype drive the specificity in the interaction of a natural host-parasite system. *Ecology letters* 15**,** 1095-1103.

Kojima, Y., Toki, T., Morimoto, T., Yoshiyama, M., Kimura, K., and Kadowaki, T. (2011). Infestation of Japanese native honey bees by tracheal mite and virus from non-native European honey bees in Japan. *Microb Ecol* 62**,** 895-906.

Kolev, N.G., Tschudi, C., and Ullu, E. (2011). RNA interference in protozoan parasites: achievements and challenges. *Eukaryotic cell***,** EC. 05114-05111.

Kondo, S., and Ueda, R. (2013). Highly improved gene targeting by germline-specific Cas9 expression in Drosophila. *Genetics***,** genetics. 113.156737.

Kulkarni, M.M., Mcmaster, W.R., Kamysz, E., Kamysz, W., Engman, D.M., and Mcgwire, B.S. (2006). The major surface-metalloprotease of the parasitic protozoan, Leishmania, protects against antimicrobial peptide-induced apoptotic killing. *Molecular microbiology* 62**,** 1484-1497.

Kuraishi, T., Binggeli, O., Opota, O., Buchon, N., and Lemaitre, B. (2011). Genetic evidence for a protective role of the peritrophic matrix against intestinal bacterial infection in Drosophila melanogaster. *Proceedings of the National Academy of Sciences***,** 201105994.

Kwong, W.K., and Moran, N.A. (2013). Cultivation and characterization of the gut symbionts of honey bees and bumble bees: description of Snodgrassella alvi gen. nov., sp. nov., a member of the family Neisseriaceae of the Betaproteobacteria, and Gilliamella apicola gen. nov., sp. nov., a member of Orbaceae fam. nov., Orbales ord. nov., a sister taxon to the order 'Enterobacteriales' of the Gammaproteobacteria. *Int J Syst Evol Microbiol* 63**,** 2008-2018.

Kwong, W.K., and Moran, N.A. (2015). Evolution of host specialization in gut microbes: the bee gut as a model. *Gut Microbes* 6**,** 214-220.

Kwong, W.K., and Moran, N.A. (2016). Gut microbial communities of social bees. *Nature Reviews Microbiology* 14**,** 374.

Lacount, D.J., Barrett, B., and Donelson, J.E. (2002). Trypanosoma brucei FLA1 is required for flagellum attachment and cytokinesis. *J Biol Chem* 277**,** 17580-17588.

Lander, N., Chiurillo, M.A., and Docampo, R. (2016a). Genome editing by CRISPR/Cas9: a game change in the genetic manipulation of protists. *Journal of Eukaryotic Microbiology* 63**,** 679-690.

Lander, N., Chiurillo, M.A., Storey, M., Vercesi, A.E., and Docampo, R. (2016b). CRISPR/Cas9-mediated endogenous C-terminal tagging of Trypanosoma cruzi genes reveals the acidocalcisome localization of the inositol 1, 4, 5-trisphosphate receptor. *Journal of Biological Chemistry* 291**,** 25505-25515.

Lander, N., Chiurillo, M.A., Vercesi, A.E., and Docampo, R. (2017). Endogenous C-terminal tagging by CRISPR/Cas9 in Trypanosoma cruzi. *Bio-protocol* 7.

Lander, N., Li, Z.-H., Niyogi, S., and Docampo, R. (2015a). CRISPR/Cas9-induced disruption of paraflagellar rod protein 1 and 2 genes in Trypanosoma cruzi reveals their role in flagellar attachment. *MBio* 6**,** e01012-01015.

Lander, N., Li, Z.H., Niyogi, S., and Docampo, R. (2015b). CRISPR/Cas9-Induced Disruption of Paraflagellar Rod Protein 1 and 2 Genes in Trypanosoma cruzi Reveals Their Role in Flagellar Attachment. *MBio* 6**,** e01012.

Langridge, D.F., and Mcghee, R.B. (1967). Crithidia mellificae n. sp. an acidophilic trypanosomatid of the honey bee Apis mellifera. *J Protozool* 14**,** 485-487.

Lee, F.J., Rusch, D.B., Stewart, F.J., Mattila, H.R., and Newton, I.L. (2015). Saccharide breakdown and fermentation by the honey bee gut microbiome. *Environmental microbiology* 17**,** 796-815.

Lehane, M. (1997). Peritrophic matrix structure and function. *Annual review of entomology* 42**,** 525-550.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25**,** 2078-2079.

Li, J.H., Evans, J.D., Li, W.F., Zhao, Y.Z., Degrandi-Hoffman, G., Huang, S.K., Li, Z.G., Hamilton, M., and Chen, Y.P. (2017). New evidence showing that the destruction of gut bacteria by antibiotic treatment could increase the honey bee's vulnerability to Nosema infection. *PloS one* 12**,** e0187505.

Loch, G., Zinke, I., Mori, T., Carrera, P., Schroer, J., Takeyama, H., and Hoch, M. (2017). Antimicrobial peptides extend lifespan in Drosophila. *PloS one* 12**,** e0176689.

Lourenço, A.P., Guidugli-Lazzarini, K.R., Freitas, F.C., Bitondi, M.M., and Simões, Z.L. (2013). Bacterial infection activates the immune system response and dysregulates microRNA expression in honey bees. *Insect biochemistry and molecular biology* 43**,** 474-482.

Ludvigsen, J., Rangberg, A., Avershina, E., Sekelja, M., Kreibich, C., Amdam, G., and Rudi, K. (2015). Shifts in the midgut/pyloric microbiota composition within a honey bee apiary throughout a season. *Microbes and environments* 30**,** 235-244.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., and Gonzales, N.R. (2016). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic acids research* 45**,** D200-D203.

Marchler-Bauer, A., and Bryant, S.H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic acids research* 32**,** W327-W331.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., and Hurwitz, D.I. (2014). CDD: NCBI's conserved domain database. *Nucleic acids research* 43**,** D222-D226.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., Deweese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., and Gonzales, N.R. (2010). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic acids research* 39**,** D225-D229.

Martel, D., Beneke, T., Gluenz, E., Späth, G.F., and Rachidi, N. (2017). Characterisation of Casein Kinase 1.1 in Leishmania donovani Using the CRISPR Cas9 Toolkit. *BioMed research international* 2017.

Martin, J.-R. (2003). Locomotor activity: a complex behavioural trait to unravel. *Behavioural Processes* 64**,** 145-160.

Martin, J.-R. (2004). A portrait of locomotor behaviour in Drosophila determined by a video-tracking paradigm. *Behavioural processes* 67**,** 207-219.

Martínez-Calvillo, S., Vizuet-De-Rueda, J.C., Florencio-Martínez, L.E., Manning-Cela, R.G., and Figueroa-Angulo, E.E. (2010). Gene expression in trypanosomatid parasites. *BioMed Research International* 2010.

MartíNez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K., and Myler, P.J. (2003). Transcription of Leishmania major Friedlin chromosome 1 initiates in both directions within a single region. *Molecular cell* 11**,** 1291-1299.

Martinson, V.G., Danforth, B.N., Minckley, R.L., Rueppell, O., Tingek, S., and Moran, N.A. (2011). A simple and distinctive microbiota associated with honey bees and bumble bees. *Mol Ecol* 20**,** 619-628.

Martinson, V.G., Moy, J., and Moran, N.A. (2012). Establishment of characteristic gut bacteria during development of the honey bee worker. *Applied and environmental microbiology***,** AEM. 07810-07811.

Mayans, O., Benian, G.M., Simkovic, F., and Rigden, D.J. (2013). "Mechanistic and functional diversity in the mechanosensory kinases of the titin-like family". Portland Press Limited).

Mellroth, P., Karlsson, J., Håkansson, J., Schultz, N., Goldman, W.E., and Steiner, H. (2005). Ligand-induced dimerization of Drosophila peptidoglycan recognition proteins in vitro. *Proceedings of the National Academy of Sciences* 102**,** 6455-6460.

Meunier, N., Belgacem, Y.H., and Martin, J.-R. (2007). Regulation of feeding behaviour and locomotor activity by takeout in Drosophila. *Journal of Experimental Biology* 210**,** 1424-1434.

Michaeli, S. (2011). Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future microbiology* 6**,** 459-474.

Mockler, B.K., Kwong, W.K., Moran, N.A., and Koch, H. (2018). Microbiome structure influences infection by the parasite Crithidia bombi in bumble bees. *Applied and environmental microbiology***,** AEM. 02335-02317.

Moran, N.A., Hansen, A.K., Powell, J.E., and Sabree, Z.L. (2012). Distinctive gut microbiota of honey bees assessed using deep sampling from individual worker bees. *PloS one* 7**,** e36393.

Moreno, M.A., Alonso, A., Alcolea, P.J., Abramov, A., De Lacoba, M.G., Abendroth, J., Zhang, S., Edwards, T., Lorimer, D., and Myler, P.J. (2014). Tyrosine aminotransferase from Leishmania infantum: A new drug target candidate. *International Journal for Parasitology: Drugs and Drug Resistance* 4**,** 347-354.

Morimoto, T., Kojima, Y., Yoshiyama, M., Kimura, K., Yang, B., Peng, G., and Kadowaki, T. (2013). Molecular detection of protozoan parasites infecting Apis mellifera colonies in Japan. *Environmental microbiology reports* 5**,** 74-77.

Motyka, S.A., Drew, M.E., Yildirir, G., and Englund, P.T. (2006). Over-expression of a cytochrome b5 reductase-like protein causes kinetoplast DNA loss in Trypanosoma brucei. *Journal of Biological Chemistry*.

Müller, U. (1996). Inhibition of nitric oxide synthase impairs a distinct form of long-term memory in the honeybee, Apis mellifera. *Neuron* 16**,** 541-549.

Müller, U., and Hildebrandt, H. (2002). Nitric oxide/cGMP-mediated protein kinase A activation in the antennal lobes plays an important role in appetitive reflex habituation in the honeybee. *Journal of Neuroscience* 22**,** 8739-8747.

Myler, P.J., Audleman, L., Hixson, G., Kiser, P., Lemley, C., Magness, C., Rickel, E., Sisk, E., Sunkin, S., and Swartzell, S. (1999). Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proceedings of the National Academy of Sciences* 96**,** 2902-2906.

Negri, P., Maggi, M., Correa-Aragunde, N., Brasesco, C., Eguaras, M., and Lamattina, L. (2013). Nitric oxide participates at the first steps of Apis mellifera cellular immune activation in response to non-self recognition. *Apidologie* 44**,** 575-585.

Nelson, C.M., Ihle, K.E., Fondrk, M.K., Page Jr, R.E., and Amdam, G.V. (2007). The gene vitellogenin has multiple coordinating effects on social organization. *PLoS biology* 5**,** e62.

Ngo, H., Tschudi, C., Gull, K., and Ullu, E. (1998). Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. *Proc Natl Acad Sci U S A* 95**,** 14687-14692.

Olivier, M., Atayde, V.D., Isnard, A., Hassani, K., and Shio, M.T. (2012). Leishmania virulence factors: focus on the metalloprotease GP63. *Microbes and infection* 14**,** 1377-1389.

Onn, I., Milman-Shtepel, N., and Shlomai, J. (2004). Redox potential regulates binding of universal minicircle sequence binding protein at the kinetoplast DNA replication origin. *Eukaryotic cell* 3**,** 277-287.

Oshlack, A., Robinson, M.D., and Young, M.D. (2010). From RNA-seq reads to differential expression results. *Genome biology* 11**,** 220.

Palmer-Young, E.C., Tozkar, C.Ö., Schwarz, R.S., Chen, Y., Irwin, R.E., Adler, L.S., and Evans, J.D. (2017). Nectar and pollen phytochemicals stimulate honey bee (Hymenoptera: Apidae) immunity to viral infection. *Journal of economic entomology* 110**,** 1959-1972.

Passos-Silva, D.G., Rajão, M.A., Nascimento De Aguiar, P.H., Vieira-Da-Rocha, J.P., Machado, C.R., and Furtado, C. (2010). Overview of DNA Repair in Trypanosoma cruzi, Trypanosoma brucei, and Leishmania major. *Journal of nucleic acids* 2010.

Peng, D., Kurup, S.P., Yao, P.Y., Minning, T.A., and Tarleton, R.L. (2015). CRISPR-Cas9-mediated single-gene and gene family disruption in Trypanosoma cruzi. *MBio* 6**,** e02097-02014.

Pereira, F.M., Dias, F.A., Elias, C.G., D'avila-Levy, C.M., Silva, C.S., Santos-Mallet, J.R., Branquinha, M.H., and Santos, A.L. (2010). Leishmanolysin-like molecules in Herpetomonas samuelpessoai mediate hydrolysis of protein substrates and interaction with insect. *Protist* 161**,** 589-602.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature protocols* 11**,** 1650.

Peters, W. (2012). *Peritrophic membranes.* Springer Science & Business Media.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O., and Kunin, W.E. (2010). Global pollinator declines: trends, impacts and drivers. *Trends in ecology & evolution* 25**,** 345-353.

Powell, J.E., Martinson, V.G., Urban-Mead, K., and Moran, N.A. (2014). Routes of acquisition of the gut microbiota of Apis mellifera. *Applied and environmental microbiology***,** AEM. 01861-01814.

Proudfoot, C., and Mcculloch, R. (2006). Trypanosoma brucei DMC1 does not act in DNA recombination, repair or antigenic variation in bloodstream stage cells. *Molecular and biochemical parasitology* 145**,** 245-253.

Quijada, L., Guerra-Giraldez, C., Drozdz, M., Hartmann, C., Irmer, H., Ben-Dov, C., Cristodero, M., Ding, M., and Clayton, C. (2002). Expression of the human RNA-binding protein HuR in Trypanosoma brucei increases the abundance of mRNAs containing AU-rich regulatory elements. *Nucleic acids research* 30**,** 4414-4424.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature protocols* 8**,** 2281-2308.

Ravoet, J., De Smet, L., Meeus, I., Smagghe, G., Wenseleers, T., and De Graaf, D.C. (2014). Widespread occurrence of honey bee pathogens in solitary bees. *Journal of Invertebrate Pathology* 122**,** 55-58.

Ravoet, J., Maharramov, J., Meeus, I., De Smet, L., Wenseleers, T., Smagghe, G., and De Graaf, D.C. (2013). Comprehensive bee pathogen screening in Belgium reveals Crithidia mellificae as a new contributory factor to winter mortality. *PLoS One* 8**,** e72443.

Ravoet, J., Schwarz, R.S., Descamps, T., Yanez, O., Tozkar, C.O., Martin-Hernandez, R., Bartolome, C., De Smet, L., Higes, M., Wenseleers, T., Schmid-Hempel, R., Neumann, P., Kadowaki, T., Evans, J.D., and De Graaf, D.C. (2015). Differential diagnosis of the honey bee trypanosomatids Crithidia mellificae and Lotmaria passim. *J Invertebr Pathol* 130**,** 21-27.

Raymann, K., Shaffer, Z., and Moran, N.A. (2017). Antibiotic exposure perturbs the gut microbiota and elevates mortality in honeybees. *PLoS biology* 15**,** e2001861.

Regan, T., Barnett, M.W., Laetsch, D.R., Bush, S.J., Wragg, D., Budge, G.E., Highet, F., Dainat, B., De Miranda, J.R., and Blaxter, M. (2018). Characterisation of the UK honey bee (Apis mellifera) metagenome. *bioRxiv***,** 293647.

Richardson, L.L., Adler, L.S., Leonard, A.S., Andicoechea, J., Regan, K.H., Anthony, W.E., Manson, J.S., and Irwin, R.E. (2015). Secondary metabolites in floral nectar reduce parasite infections in bumblebees. *Proc. R. Soc. B* 282**,** 20142471.

Rico, E., Jeacock, L., Kovářová, J., and Horn, D. (2018). Inducible high-efficiency CRISPR-Cas9-targeted gene editing and precision base editing in African trypanosomes. *Scientific Reports* 8**,** 7960.

Riddell, C., Adams, S., Schmid-Hempel, P., and Mallon, E.B. (2009). Differential expression of immune defences is associated with specific host-parasite interactions in insects. *PLoS One* 4**,** e7621.

Riddell, C., Sumner, S., Adams, S., and Mallon, E. (2011). Pathways to immunity: temporal dynamics of the bumblebee (Bombus terrestris) immune response against a trypanosomal gut parasite. *Insect molecular biology* 20**,** 529-540.

Riddell, C.E., Garces, J.D.L., Adams, S., Barribeau, S.M., Twell, D., and Mallon, E.B. (2014). Differential gene expression and alternative splicing in insect immune specificity. *BMC genomics* 15**,** 1031.

Rivero, A. (2006). Nitric oxide: an antiparasitic molecule of invertebrates. *Trends in parasitology* 22**,** 219-225.

Robinson, K.A., and Beverley, S.M. (2003). Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite Leishmania. *Mol Biochem Parasitol* 128**,** 217-228.

Robinson, M.D., Mccarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26**,** 139-140.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology* 11**,** R25.
Roulston, T.a.H., and Cane, J.H. (2000). Pollen nutritional content and digestibility for animals. *Plant systematics and Evolution* 222**,** 187-209.
Runckel, C., Derisi, J., and Flenniken, M.L. (2014). A draft genome of the honey bee trypanosomatid parasite Crithidia mellificae. *PLoS One* 9**,** e95057.
Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16**,** 944-945.
Ryu, J.-H., Kim, S.-H., Lee, H.-Y., Bai, J.Y., Nam, Y.-D., Bae, J.-W., Lee, D.G., Shin, S.C., Ha, E.-M., and Lee, W.-J. (2008). Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in Drosophila. *science* 319**,** 777-782.
Salathe, R., Tognazzo, M., Schmid-Hempel, R., and Schmid-Hempel, P. (2012). Probing mixed-genotype infections I: extraction and cloning of infections from hosts of the trypanosomatid Crithidia bombi. *PLoS One* 7**,** e49046.
Sant'anna, M.R., Diaz-Albiter, H., Aguiar-Martins, K., Al Salem, W.S., Cavalcante, R.R., Dillon, V.M., Bates, P.A., Genta, F.A., and Dillon, R.J. (2014). Colonisation resistance in the sand fly gut: Leishmania protects Lutzomyia longipalpis from bacterial infection. *Parasit Vectors* 7**,** 329.
Santrich, C., Moore, L., Sherwin, T., Bastin, P., Brokaw, C., Gull, K., and Lebowitz, J. (1997). A motility function for the paraflagellar rod of Leishmania parasites revealed by PFR-2 gene knockouts. *Molecular and biochemical parasitology* 90**,** 95-109.
Sarov-Blat, L., So, W.V., Liu, L., and Rosbash, M. (2000). The Drosophila takeout gene is a novel molecular link between circadian rhythms and feeding behavior. *Cell* 101**,** 647-656.
Schlüns, H., and Crozier, R. (2007). Relish regulates expression of antimicrobial peptide genes in the honeybee, Apis mellifera, shown by RNA interference. *Insect molecular biology* 16**,** 753-759.
Schlüns, H., Sadd, B.M., Schmid-Hempel, P., and Crozier, R.H. (2010). Infection with the trypanosome Crithidia bombi and expression of immune-related genes in the bumblebee Bombus terrestris. *Developmental & Comparative Immunology* 34**,** 705-709.
Schmid-Hempel, R., and Tognazzo, M. (2010). Molecular divergence defines two distinct lineages of Crithidia bombi (Trypanosomatidae), parasites of bumblebees. *Journal of Eukaryotic Microbiology* 57**,** 337-345.
Schmidt, K., and Engel, P. (2016). Probiotic treatment with a gut symbiont leads to parasite susceptibility in honey bees. *Trends in parasitology* 32**,** 914-916.
Schwarz, R.S., Bauchan, G.R., Murphy, C.A., Ravoet, J., De Graaf, D.C., and Evans, J.D. (2015). Characterization of Two Species of Trypanosomatidae from the Honey Bee Apis mellifera: Crithidia mellificae Langridge and McGhee, and Lotmaria passim n. gen., n. sp. *J Eukaryot Microbiol* 62**,** 567-583.
Schwarz, R.S., and Evans, J.D. (2013). Single and mixed-species trypanosome and microsporidia infections elicit distinct, ephemeral cellular and humoral immune responses in honey bees. *Developmental & Comparative Immunology* 40**,** 300-310.
Schwarz, R.S., Moran, N.A., and Evans, J.D. (2016). Early gut colonizers shape parasite susceptibility and microbiota composition in honey bee workers. *Proc Natl Acad Sci U S A* 113**,** 9345-9350.

Seehuus, S.-C., Norberg, K., Gimsa, U., Krekling, T., and Amdam, G.V. (2006). Reproductive protein protects functionally sterile honey bee workers from oxidative stress. *Proceedings of the National Academy of Sciences* 103**,** 962-967.

Sela, D., Yaffe, N., and Shlomai, J. (2008). Enzymatic mechanism controls redox-mediated protein-DNA interactions at the replication origin of kinetoplast DNA minicircles. *J Biol Chem* 283**,** 32034-32044.

Seol, J.-H., Shim, E.Y., and Lee, S.E. (2017). Microhomology-mediated end joining: good, bad and ugly. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*.

Sfeir, A., and Symington, L.S. (2015). Microhomology-mediated end joining: a back-up survival mechanism or dedicated pathway? *Trends in biochemical sciences* 40**,** 701-714.

Shanbhag, S., Hekmat-Scafe, D., Kim, M.S., Park, S.K., Carlson, J., Pikielny, C., Smith, D., and Steinbrecht, R. (2001). Expression mosaic of odorant-binding proteins in Drosophila olfactory organs. *Microscopy research and technique* 55**,** 297-306.

Shao, L., Devenport, M., and Jacobs-Lorena, M. (2001). The peritrophic matrix of hematophagous insects. *Archives of insect biochemistry and physiology* 47**,** 119-125.

Sharma, S., Javadekar, S., Pandey, M., Srivastava, M., Kumari, R., and Raghavan, S. (2016). Homology and enzymatic requirements of microhomology-dependent alternative end joining. *Cell death & disease* 6**,** e1697.

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., Huang, X., and Skarnes, W.C. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods* 11**,** 399-402.

Shio, M.T., and Olivier, M. (2010). Leishmania survival mechanisms: the role of host phosphatases. *Journal of leukocyte biology* 88**,** 1-3.

Shlomai, J. (2010). Redox control of protein-DNA interactions: from molecular mechanisms to significance in signal transduction, gene expression, and DNA replication. *Antioxid Redox Signal* 13**,** 1429-1476.

Shykoff, J.A., and Schmid-Hempel, P. (1991). Parasites delay worker reproduction in bumblebees: consequences for eusociality. *Behavioral Ecology* 2**,** 242-248.

Singh, J., Srivastava, A., Jha, P., Sinha, K.K., and Kundu, B. (2015). L-Asparaginase as a new molecular target against leishmaniasis: insights into the mechanism of action and structure-based inhibitor design. *Mol Biosyst* 11**,** 1887-1896.

Singh, R., Purkait, B., Abhishek, K., Saini, S., Das, S., Verma, S., Mandal, A., Ghosh, A.K., Ansari, Y., Kumar, A., Sardar, A.H., Kumar, A., Parrack, P., and Das, P. (2016). Universal minicircle sequence binding protein of Leishmania donovani regulates pathogenicity by controlling expression of cytochrome-b. *Cell Biosci* 6**,** 13.

Sollelis, L., Ghorbal, M., Macpherson, C.R., Martins, R.M., Kuk, N., Crobu, L., Bastien, P., Scherf, A., Lopez-Rubio, J.J., and Sterkers, Y. (2015a). First efficient CRISPR-Cas9-mediated genome editing in Leishmania parasites. *Cell Microbiol* 17**,** 1405-1412.

Sollelis, L., Ghorbal, M., Macpherson, C.R., Martins, R.M., Kuk, N., Crobu, L., Bastien, P., Scherf, A., Lopez-Rubio, J.J., and Sterkers, Y. (2015b). First efficient CRISPR-Cas9-mediated genome editing in Leishmania parasites. *Cellular microbiology* 17**,** 1405-1412.

Steel, R.G.D., and Torrie, J.H. (1960). Principles and procedures of statistics. *Principles and procedures of statistics.*

Steinbiss, S., Silva-Franco, F., Brunk, B., Foth, B., Hertz-Fowler, C., Berriman, M., and Otto, T.D. (2016). Companion: a web server for annotation and analysis of parasite genomes. *Nucleic Acids Res* 44**,** W29-34.

Stevanovic, J., Schwarz, R.S., Vejnovic, B., Evans, J.D., Irwin, R.E., Glavinic, U., and Stanimirovic, Z. (2016). Species-specific diagnostics of Apis mellifera trypanosomatids: a nine-year survey (2007–2015) for trypanosomatids and microsporidians in Serbian honey bees. *Journal of invertebrate pathology* 139**,** 6-11.

Szalanski, A., Trammel, C., Tripodi, A., Cleary, D., Rusert, L., and Downey, D. (2016). Molecular diagnostics of the honey bee parasites Lotmaria passim and Crithidia spp.(Trypanosomatidae) using multiplex PCR. *Florida Entomologist* 99**,** 793-795.

Tanji, T., Hu, X., Weber, A.N., and Ip, Y.T. (2007). Toll and IMD pathways synergistically activate an innate immune response in Drosophila melanogaster. *Molecular and cellular biology* 27**,** 4578-4588.

Tarpy, D.R., Mattila, H.R., and Newton, I.L. (2015). Characterization of the honey bee microbiome throughout the queen-rearing process. *Applied and environmental microbiology***,** AEM. 00307-00315.

Teixeira, S. (1998). Control of gene expression in Trypanosomatidae. *Brazilian journal of medical and biological research* 31**,** 1503-1516.

Tellam, R. (1996). "The peritrophic matrix," in *Biology of the insect midgut*. Springer), 86-114.

Tellam, R.L., Wijffels, G., and Willadsen, P. (1999). Peritrophic matrix proteins. *Insect biochemistry and molecular biology* 29**,** 87-101.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25**,** 1105-1111.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7**,** 562-578.

Tritschler, M., Retschnig, G., Yañez, O., Williams, G.R., and Neumann, P. (2017). Host sharing by the honey bee parasites Lotmaria passim and Nosema ceranae. *Ecology and evolution* 7**,** 1850-1857.

Tsuzuki, S., Ochiai, M., Matsumoto, H., Kurata, S., Ohnishi, A., and Hayakawa, Y. (2012). Drosophila growth-blocking peptide-like factor mediates acute immune reactions during infectious and non-infectious stress. *Scientific reports* 2**,** 210.

Tzou, P., Reichhart, J.-M., and Lemaitre, B. (2002). Constitutive expression of a single antimicrobial peptide can restore wild-type resistance to infection in immunodeficient Drosophila mutants. *Proceedings of the National Academy of Sciences* 99**,** 2152-2157.

Vavilova, V., Konopatskaia, I., Luzyanin, S., Woyciechowski, M., and Blinov, A. (2017). Crithidia and Lotmaria in the honeybee and bumblebee populations: a case study in india. *Вавиловский журнал генетики и селекции* 21**,** 943-951.

Vazquez, M.P., and Levin, M.J. (1999). Functional analysis of the intergenic regions of TcP2beta gene loci allowed the construction of an improved Trypanosoma cruzi expression vector. *Gene* 239**,** 217-225.

Vejnovic, B., Stevanovic, J., Schwarz, R.S., Aleksic, N., Mirilovic, M., Jovanovic, N.M., and Stanimirovic, Z. (2018). Quantitative PCR assessment of Lotmaria passim in Apis mellifera colonies co-infected naturally with Nosema ceranae. *Journal of invertebrate pathology* 151**,** 76-81.

Villarreal, D.D., Lee, K., Deem, A., Shim, E.Y., Malkova, A., and Lee, S.E. (2012). Microhomology directs diverse DNA break repair pathways and chromosomal translocations. *PLoS genetics* 8**,** e1003026.

Wang, H., and Xu, X. (2017). Microhomology-mediated end joining: new players join the team. *Cell & bioscience* 7**,** 6.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153**,** 910-918.

Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., Cayouette, M., Mchardy, A.C., Djordjevic, G., and Aboushadi, N. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450**,** 560.

Weinman, D., and Cheong, W. (1978). Herpetomonas, with Bacterium-like Inclusions, in Malaysian Aedes aegypti and Aedes albopictus. *The Journal of Protozoology* 25**,** 167-169.

Werner, T., Borge-Renberg, K., Mellroth, P., Steiner, H., and Hultmark, D. (2003). Functional diversity of the Drosophila PGRP-LC gene cluster in the response to lipopolysaccharide and peptidoglycan. *Journal of Biological Chemistry* 278**,** 26319-26322.

Wheeler, D.E., and Kawooya, J.K. (1990). Purification and characterization of honey bee vitellogenin. *Archives of insect biochemistry and physiology* 14**,** 253-267.

Williams, G.R., Alaux, C., Costa, C., Csáki, T., Doublet, V., Eisenhardt, D., Fries, I., Kuhn, R., Mcmahon, D.P., and Medrzycki, P. (2013). Standard methods for maintaining adult Apis mellifera in cages under in vitro laboratory conditions. *Journal of Apicultural Research* 52**,** 1-36.

Xu, G., Palmer-Young, E., Skyrm, K., Daly, T., Sylvia, M., Averill, A., and Rich, S. (2018). Triplex real-time PCR for detection of Crithidia mellificae and Lotmaria passim in honey bees. *Parasitology research* 117**,** 623-628.

Yao, C. (2010). Major surface protease of trypanosomatids: one size fits all? *Infection and immunity* 78**,** 22-31.

Yao, C., Donelson, J.E., and Wilson, M.E. (2003). The major surface protease (MSP or GP63) of Leishmania sp. Biosynthesis, regulation of expression, and function. *Molecular and biochemical parasitology* 132**,** 1-16.

Zhang, W.-W., Lypaczewski, P., and Matlashewski, G. (2017). Optimized CRISPR-Cas9 genome editing for Leishmania and its use to target a multigene family, induce chromosomal translocation, and study DNA break repair mechanisms. *mSphere* 2**,** e00340-00316.

Zhang, W.W., and Matlashewski, G. (2015). CRISPR-Cas9-Mediated Genome Editing in Leishmania donovani. *MBio* 6**,** e00861.

Zheng, H., Powell, J.E., Steele, M.I., Dietrich, C., and Moran, N.A. (2017). Honeybee gut microbiota promotes host weight gain via bacterial metabolism and hormonal signaling. *Proceedings of the National Academy of Sciences* 114**,** 4775-4780.

7. Appendix 1 Primer list

Primers	Sequence (5'-3')	Purposes
LpMT outer 3636 F	AGACCGCGATGATGAGGTTC	Amplify outer template of *LpMT* allele
LpMT outer 3636 R	GACACCTGGTCGTGTCTTCC	
LpMT HR1 Fragment 1-F	CTACATGCACCGGTCTTTCTC	Amplify *LpMT* 5' UTR
LpMT HR1 Fragment 1-R	CTTTTTCATCGCACAGAACTACAAATT	
LpMT HygR Fragment 2-F	TTCTGTGCGATGAAAAAGCCTGAACTC	Amplify *Hyg* for constructing *LpMT* donor DNA
LpMT HygR Fragment 2-R	AATCCACCCCTATTCCTTTGCCCTCGG	
LpMT HR2 Fragment 3-F	AAGGAATAGGGGTGGATTGCCACATGA	Amplify *LpMT* 500bp ORF
LpMT HR2 Fragment 3-R	GCTCATCATCTTGCGGTCCGG	
TAT outer 2249-F	GAAGAGTACAGTGCGTCCCG	Amplify outer template of *LpMT* allele
TAT outer 2249-R	AACACAGGGCAGACTTTCGG	
TAT frag 1 F QS	GCCTTAGTTCACCTCCACGTTTCT	Amplify *Lp TAT* 5' UTR
TAT frag 1 hyg R	GGCTTTTTCATTGTGGTGACTCTTTTTGT	
TAT hyg-F	GTCACCACAATGAAAAAGCCTGAACTC	Amplify *Hyg* for constructing *LpTAT* donor DNA
TAT hyg-R	AAAGAACCGCCCTATTCCTTTGCCCTCGGA	
TAT frag 3 hyg F	AAGGAATAGGGCGGTTCTTTCTTTCTC	Amplify *Lp TAT* 3' UTR
TAT frag 3 R QS	CAGTAGTTCTGAAGGCTTTTCTCTCTTCC	
803-F (Hyg seq-F)	ACCGCAAGGAATCGGTCAAT	Detect *LpMT* KO allele
803-R (*LpMT*-500 F)	AGGTCATTCGTGTCGATCGGGA	
1230-F (*LpMT* HR1 Fragment 1-F)	CTACATGCACCGGTCTTTCTC	Other primers to detect *LpMT* KO allele
1230-R (Hyg seq-R)	GGCGACCTCGTATTGGGAAT	
851-F (Homo test 851-F)	TACGGCTCTTCAGGCTTGTC	Detect *LpMT* WT allele
851-R (Homo test 851-R)	ATGCGTACAAAGGAGGGACC	
564-F (*LpMT* 564 F)	CTTGAAGGGCTGGATCGTGT	Run for *LpMT* RT-PCR
564-R (*LpMT* 564 R)	AACCTCGACCGCAAGCTAAA	
TAT 455 F	GCAACTATAGCCGCAAGCAC	Run for *LpTAT* RT-PCR
TAT 455 R	CATGACGAACATCGACGCAC	
L. passim GAPDH-F	TGAACGGCCACCGCATCCTG	Run for *GAPDH* RT-PCR (Kojima et al., 2011)
L. passim GPADH-R	GGGCCAGGCAGTTGGTCGTG	
1128-F (TAT frag 1 F QS)	GCCTTAGTTCACCTCCACGTTTCT	Detect *Lp TAT* KO allele
1128-R (Hyg seq-R)	GGCGACCTCGTATTGGGAAT	
1143-F (TAT nest 1443-F)	TCGACCGTTTTTCTCGCCAT	Detect *Lp TAT* WT allele
1143-R (TAT 455 R)	CATGACGAACATCGACGCAC	
LpMT 500-F	AGGTCATTCGTGTCGATCGGGA	T7E1 treatment *LpMT* allele
LpMT 500-R	ACGAAGATGTTCCGCAACTTGAAGG	
TAT 1106-F(TAT nest 1443-F)	TCGACCGTTTTTCTCGCCAT	T7E1 treatment *LpTAT* allele
TAT 1106-R(TAT 455 R)	CATGACGAACATCGACGCAC	
Lp-qPCR 2nd-72 F	GGGTCTTTTGTGATCGGGATAA	Detect the relative abundance of *L. passim* by qPCR at 5 time points
Lp-qPCR 2nd-72 R	CAAAAAGATGCCTAACGTGAAGAA	
AmHsTRPA-767	TAGCGTACATGTGGTGCTGT	reference gene in qPCR to detecting the relative abundance of *L. passim* at

AmHsTRPA-858	GCTAGGCTCCACGTAATCCA	5 time points
Honey bee Vg-F	GTTGGAGAGCAACATGCAGA	Analyze honey bee vitellogenin in the fat body using qPCR(Ihle et al., 2015)
Honey bee Vg-R	TCGATCCATTCCTTGATGGT	
Bee 18s rRNA-416	ACCACATCCAAGGAAGGCAG	endogenous control for Vg. Also used as reference gene for RNA-seq samples selection
Bee 18s rRNA-527	ACTCATTCCGATTACGGGGC	
Lp 18S RNA-87 F	GAAAGGAACCACTCCCGTGT	Analyze the relative abundance of *L. passim* for RNA-seq samples selection
Lp 18S RNA-87 R	GTCCCGTCCATGTCGGATTT	
Univ bacteria 328-F	AGAGTTTGATCCTGGCTCAG	Ananlyze the universal bacteria in the gut in response to *L. passim* infection (Powell et al., 2014)
Univ bacteria 328-R	CTGCTGCCTCCCGTAGGAGT	
Firmicutes-F	TGAAACTYAAAGGAATTGACG	Ananlyze the *Firmicutes* in the gut in response to *L. passim* infection (De Gregoris et al., 2011)
Firmicutes-R	ACCATGCACCACCTGTC	
Apis-β-actin-F	TTGTATGCCAACACTGTCCTTT	reference gene in qPCR to detecting universal bacteria (Li et al., 2017)
Apis-β-actin-R	TGGCGCGATGATCTTAATTT	
L. passim ITS1-IR1	GCTGTAGGTGAACCTGCAGCAGCTGGATCATT	Detect *L. passim* from mortality test (Maia Da Silva et al., 2004)
L. passim ITS1-5.8R	GGAAGCCAAGTCATCCATC	
Honey bee *AmHsTRPA*-F	CACGACATTCAAGGTTTAAGAAATCACG	use as control of honey bee in the mortality test (Kojima et al., 2011)
Honey bee *AmHsTRPA*-R	TCAGTTATTCTTTTCCTTTGCCAGATTT	
TAT sgRNA 111-F	TTGTGATCCCTGATCAAGCTGTCGA	sgRNA targeting to *LpTAT* on sgRNA expression vector
TAT sgRNA 111-R	AAACTCGACAGCTTGATCAGGGATC	
sgRNA-Cas9-F (BbsI)	TTGTGGACAAGCTGTTCATCCAGC	sgRNA targeting to Cas9 ORF on sgRNA expression vector
sgRNA-Cas9-R (BbsI)	AAACGCTGGATGAACAGCTTGTCC	
sgRNA-*LpMT*-F (BbsI)	TTGTATTCAGGCGCAGTGGATGC	sgRNA targeting to *LpMT* on sgRNA expression vector
sgRNA-*LpMT*-R	AAACGCATCCACTGCGCCTGAAT	
Puro cass 817 F	CGGTGTCCCTCAGGATGTTT	Detect upstream of Cas9 ORF by puromycin cassette KO
Puro cass 817 R	AGTTCTTGCAGCTCGGTGAC	
Puro cass 701 F	CGCAACCTCCCCTTCTACGA	Detect downstream of Cas9 ORF by puromycin cassette KO
Puro cass 701 R	GGTGCCGTCCATCTTTTCCA	

Kojima, Y., Toki, T., Morimoto, T., Yoshiyama, M., Kimura, K., and Kadowaki, T. (2011). Infestation of Japanese native honey bees by tracheal mite and virus from non-native European honey bees in Japan. *Microb Ecol* 62, 895-906.

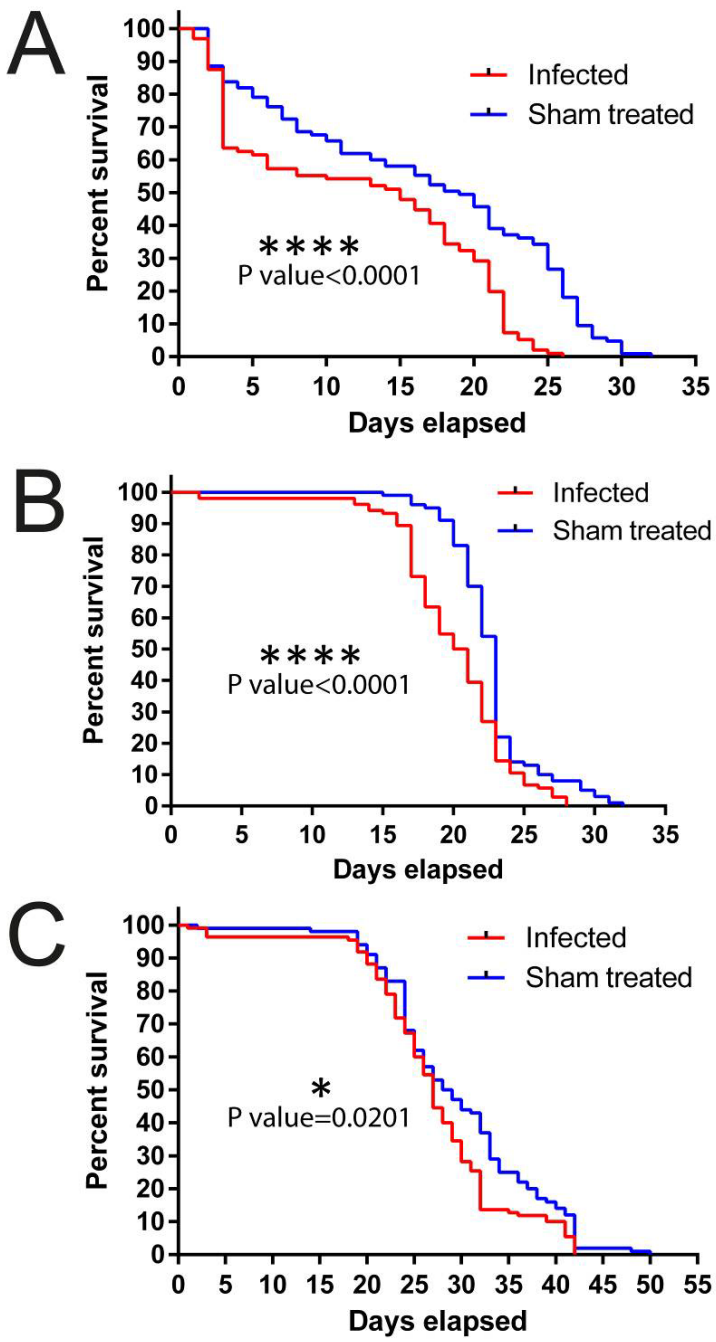
Ihle, K.E., Rueppell, O., Huang, Z.Y., Wang, Y., Fondrk, M.K., Page Jr, R.E., and Amdam, G.V. (2015). Genetic architecture of a hormonal response to gene knockdown in honey bees. *Journal of Heredity* 106, 155-165.

Maia Da Silva, F., Noyes, H., Campaner, M., Junqueira, A.C., Coura, J.R., Anez, N., Shaw, J.J., Stevens, J.R., Teixeira, M.M., 2004. Phylogeny, taxonomy and grouping of Trypanosoma rangeli isolates from man, triatomines and sylvatic mammals from widespread geographical origin based on SSU and ITS ribosomal sequences. Parasitology 129, 549-561.

De Gregoris, T.B., Aldred, N., Clare, A.S., and Burgess, J.G. (2011). Improvement of phylum-and class-specific primers for real-time PCR quantification of bacterial c. *Journal of microbiological methods* 86, 351-356.

Li, J.H., Evans, J.D., Li, W.F., Zhao, Y.Z., Degrandi-Hoffman, G., Huang, S.K., Li, Z.G., Hamilton, M., and Chen, Y.P. (2017). New evidence showing that the destruction of gut bacteria by antibiotic treatment could increase the honey bee's vulnerability to Nosema infection. *PloS one* 12, e0187505.

8. Appendix 2 Individual mortality test at 33 degree under lab condition



Three individual mortality tests from *L. passim* infected group and sham treated group. The data were analyzed by Log-rank (Mantel-Cox) test. *P-value*s were shown each group.

9. Appendix 3 HISAT2 alignment statistics

For *L. passim*

Sample ID	All reads	Aligned Mates Concordantly exactly 1 time	Aligned Mates Concordantly > 1 times	Aligned Mates Discordantly 1 time	One Mate Aligned exactly 1 time	One Mate Aligned > 1 time	Total Reads Aligned	Overall Alignment Rate	edgeR library size
INF_PI_12_A_56	16,610,073	8,960,044	1,184,429	251,361	1,952,178	450,503	11,597,153	69.82%	6,127,705
INF_PI_12_F_94	28,687,610	15,624,501	2,098,127	318,136	3,323,960	864,872	20,135,833	70.19%	9,744,902
INF_PI_20_A_95	26,271,537	15,007,553	1,932,308	596,182	4,566,890	1,121,143	20,378,831	77.57%	10,314,435
INF_PI_20_C_96	31,305,329	17,035,065	2,092,964	603,836	4,863,826	1,043,355	22,686,972	72.47%	11,337,700
INF_PI_27_D_100	25,470,882	13,815,501	1,748,341	345,929	3,617,448	832,383	18,135,268	71.20%	9,293,620
INF_PI_27_E_101	29,104,522	16,778,564	2,110,599	540,090	4,750,513	1,103,648	22,355,183	76.81%	11,468,911
YBPI7INF_210	22,271,512	2,771,827	459,383	26,223	365,850	69,893	3,474,356	15.60%	1,748,949
YF2PI7_211	21,778,943	2,673,986	466,352	31,449	396,105	79,497	3,410,582	15.66%	1,684,230
L_passim_WT_A	15,624,298	6,927,977	1,236,613	143,405	4,004,613	996,580	10,808,889	69.18%	4,658,759
L_passim_WT_B	15,713,265	8,211,470	1,243,094	186,435	3,537,291	824,606	11,822,661	75.24%	5,498,981

For honey bee

Sample ID	All reads	Aligned Mates Concordantly exactly 1 time	Aligned Mates Concordantly > 1 times	Aligned Mates Discordantly 1 time	One Mate Aligned exactly 1 time	One Mate Aligned > 1 time	Total Reads Aligned	Overall Alignment Rate	edgeR library size
CON_PI_12_E_16	14,038,354	11,575,786	195,502	368,415	1,438,815	46,326	12,881,594	91.76%	9,307,452
CON_PI_12_F_17	13,646,519	11,165,129	204,891	369,874	1,308,783	44,026	12,416,968	90.99%	8,376,520
CON_PI_20_A_20	13,933,427	11,398,942	187,260	352,531	1,492,980	56,443	12,712,859	91.24%	8,732,769
CON_PI_20_B_27	13,658,291	10,847,662	171,436	378,525	1,542,237	54,974	12,196,854	89.30%	8,351,825
CON_PI_27_A_19	13,936,795	10,642,156	161,820	279,004	1,379,197	45,350	11,794,710	84.63%	8,067,828
CON_PI_27_A_29	13,867,609	11,336,028	192,953	361,490	1,523,090	56,500	12,680,542	91.44%	8,936,807
BAPI7_control_212	22,002,553	16,933,345	379,612	133,439	963,174	71,319	17,962,884	81.64%	13,412,670
BBPI7_control_213	21,851,256	15,215,055	401,292	148,499	946,076	120,551	16,298,852	74.59%	11,354,077
INF_PI_12_A_56	16,610,073	1,736,107	29,467	13,859	264,742	9,253	1,916,802	11.54%	1,092,607
INF_PI_12_F_94	28,687,610	3,086,740	71,990	27,082	506,952	28,956	3,453,988	12.04%	1,106,598
INF_PI_20_A_95	26,271,537	1,780,869	27,340	31,624	594,977	12,838	2,143,757	8.16%	1,455,925
INF_PI_20_C_96	31,305,329	2,859,305	45,316	56,662	854,037	20,376	3,399,759	10.86%	2,317,085
INF_PI_27_D_100	25,470,882	3,018,410	47,944	43,631	626,214	16,118	3,430,928	13.47%	2,389,713
INF_PI_27_E_101	29,104,522	2,199,132	35,069	33,095	600,141	16,238	2,575,750	8.85%	1,777,864
YBPI7INF_210	22,271,512	14,287,867	297,629	121,439	809,637	61,387	15,142,401	67.99%	11,483,557
YF2PI7_211	21,778,943	10,853,675	204,070	116,017	729,918	47,528	11,562,441	53.09%	8,594,004

The complete DEGs lists and GO enrichment terms can be found in the supplementary documents.